



UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIAS E TECNOLOGIA

REDES DINÂMICAS



É talvez a REDE mais COMPLEXA e ENIGMÁTICA do universo.
Contém mais neurónios do que as estrelas existentes na galáxia

O presente Relatório refere-se ao programa, aos conteúdos e à organização pedagógica e didáctica da Unidade Curricular de Redes Dinâmicas integrante do corpo de unidades curriculares optativas do 3º Ciclo em Matemática do Instituto de Investigação e Formação Avançada da Universidade de Évora.

Conteúdo

1	Introdução	1
1.1	Decisivas contribuições para esta opção	1
1.1.1	- José Sousa Ramos	1
1.1.2	- Christian Mira	3
1.1.3	- Ton Coolen	3
1.1.4	- Mason Porter	4
1.2	A era da produção em larga escala de dados!	5
1.3	A matemática na vanguarda do conhecimento!	7
1.4	O século da complexidade !	9
1.5	A razão da existência da disciplina Redes Dinâmicas	10
1.6	Estudo do panorama geral dos cursos existentes	12
2	Descrição da disciplina Redes Dinâmicas	17
2.1	Ojectivos e competências	17
2.2	Funcionamento da disciplina	18
2.3	Programa Resumido	18
2.3.1	Grafos: generalidades e terminologia	18
2.3.2	Teoria espectral dos grafos.	19

2.3.3	Medidas de centralidade.	19
2.3.4	Redes aleatórias	20
2.3.5	Redes Small World	20
2.3.6	Redes Sem-Escala	20
2.3.7	Partição de grafos e detecção de comunidades	21
2.3.8	Sistemas Dinâmicos em Redes e de Redes	21
2.4	Cronograma	22
2.5	Metodologia e estratégias de ensino	22
2.6	Avaliação da Disciplina	23
2.6.1	Títulos dos trabalhos finais realizados pelos alunos	24
3	Sumários estendidos	27
3.1	Módulo 1 - Apresentação e Motivação	27
3.1.1	Motivação	27
3.1.2	Programa	27
3.1.3	Funcionamento da disciplina	27
3.1.4	Avaliação	27
3.2	Módulo 2 - Conceitos básicos	28
3.2.1	Grafos: generalidades e terminologia	28
3.2.2	Grafo como estrutura discreta finita	28
3.2.3	Tipos especiais de grafos.	29
3.2.4	Conceitos de passeios, caminhos, circuitos, ciclos e comprimento.	29
3.2.5	Grafos e conexidade	29
3.2.6	Isomorfismo de grafos:	29

3.2.7	Árvores:	29
3.2.8	Representação de um grafo	29
3.3	Módulo 3 - Introdução à teoria espectral de grafos	30
3.4	Módulo 4 - Medidas de centralidade	37
3.4.1	Limitações e Percepções	37
3.4.2	Evolução e diversidade de aplicação das métricas de centralidade	38
3.4.3	Centralidade pontual	39
3.4.4	Centralidade do grau	39
3.4.5	Medidas baseadas no conceito de entropia	39
3.4.6	Índice $\mathbf{h}$	40
3.4.7	Curvatura	41
3.4.8	Centralidade do vector próprio	42
3.4.9	Centralidade de um vértice por conectividade algébrica.	42
3.4.10	Centralidade de Katz	42
3.4.11	Classificação de páginas (page-Rank)	42
3.4.12	Coeficiente de aglomeramento	42
3.4.13	Centralidade de intermediação	43
3.4.14	Centralidade de Intermediação de Passeios Aleatórios	44
3.4.15	Centralidade de proximidade	45
3.4.16	Centralidade de eficiência	45
3.5	Medidas de centralidade no grafo	46
3.5.1	Número isoperimétrico	46
3.5.2	Assortatividade de grau	46

3.5.3	Curvatura do Grafo	47
3.6	Módulo 5 - Redes aleatórias	49
3.6.1	“Rede aleatória” de Solomonoff e Rapoport	49
3.6.2	O grafo aleatório de Erdős-Rényi	49
3.6.3	Relação entre estes modelos $G(n, m)$ e $G(n, p)$	50
3.6.4	Propriedades das redes aleatórias de Erdős-Rényi:	50
3.6.5	Gerando uma rede $\mathbf{G}(\mathbf{n}, \mathbf{p})$	51
3.6.6	Distribuição dos graus	51
3.6.7	Diâmetro e comprimento médio de um caminho geodésico e suas implicações	52
3.6.8	Qual é a aparência dos gráficos $G(n, p)$	52
3.6.9	Modelo de configuração	52
3.6.10	Propriedades dos grafos aleatórios do modelo de configuração	52
3.6.11	Gerando uma rede de modelo de configuração	52
3.7	Módulo 6 - Redes Small World	53
3.7.1	Coefficiente Agrupamento (Clustering): Local, global	53
3.7.2	Estudo das implicações sobre a transitividade nestas redes	55
3.7.3	Exemplos de redes: Facebook	56
3.7.4	Modelo de redes Watts e Strogatz (WS), (Nature 1998)	56
3.7.5	Propriedades de rede de mundo pequeno	58
3.7.6	Verificação deste fenómeno nalgumas redes	59
3.8	Módulo 7 - Redes sem-Escala	60
3.8.1	Motivação: redes internet	60
3.8.2	Lei de potência	60

3.8.3	Função de distribuição <i>livre de escala</i>	61
3.8.4	Redes <i>livre de escala</i>	61
3.8.5	Modelo de <i>Anexação Preferencial (preferential attachment)</i>	61
3.8.6	Rede de Barabási-Albert	62
3.8.7	Exemplos de aplicação deste modelo	63
3.9	Módulo 8 - Partição de grafos e detecção de comunidades	64
3.9.1	Motivação	64
3.9.2	Noções de complexidade computacional	65
3.9.3	Partição de um grafo	65
3.9.4	Métodos espectrais	66
3.9.5	Método de Fiedler	67
3.9.6	Detecção de comunidades	69
3.9.7	Tipos de Métodos	70
3.9.8	Algoritmo de Newman-Girvan	70
3.9.9	Modularidade	71
3.9.10	Optimização da modularidade através de passeios aleatórios	71
3.10	Módulo 9 - Sistemas Dinâmicos em e de Redes	74
3.10.1	Motivação	74
3.10.2	Distinção entre modelos para a dinâmica nas redes e a dinâmica das redes	75
3.10.3	Alguns sistemas dinâmicos estudados em redes.	75
3.10.4	Percolação	75
3.10.5	Contágios Biológicos	76
3.10.6	Contágios Sociais	77

3.10.7 Modelos de Votação	77
3.10.8 Acoplamento de osciladores	78
3.10.9 Sincronização completa de uma rede linear	78
3.10.10 Estudo da sincronizabilidade duma rede	80
3.10.11 Função zeta de um grafo	80
3.11 Bibliografia	81
A Módulo 1 - Apresentação e Motivação	89
A.0.1 Motivação	90
A.0.2 Programa	112
A.0.3 Funcionamento da disciplina	113
A.0.4 Avaliação	115
B Notas Introdutórias	117
B.1 Introdução	119
B.2 Definições	124
B.2.1 Tipos de grafos	124
B.2.2 A matriz de adjacência e de incidência de um grafo	130
B.2.3 Grau de um vértice	133
B.2.4 Distribuição dos graus	136
B.2.5 Correlação dos graus	139
B.2.6 Caminhos e conectividade	141
B.2.7 Distância	147
B.3 Agrupamento - clustering	150

B.3.1	Parte espectral	153
B.4	Centralidades	155
B.4.1	Medidas mais comuns	156
B.4.2	Centralidade de intermediação (“Betweenness centrality”)	156
B.4.3	Centralidade de Proximidade “Closeness centrality”)	157
B.4.4	Medidas de centralidade espectrais	158

Capítulo 1

Introdução

Este Relatório integrante das Provas de Agregação constitui uma oportunidade para fazer um balanço e reflectir sobre a experiência científica e pedagógica de leccionação de uma Unidade Curricular que foi criada de raiz pela autora do mesmo.

1.1 Decisivas contribuições para esta opção

1.1.1 - José Sousa Ramos

Começo por transcrever o início da minha tese de doutoramento que se mantém atual:

“...Em relação ao Prof. J. Sousa Ramos parece-me difícil nalgumas linhas exprimir toda a minha gratidão. Foi com ele que iniciei a aventura que este trabalho constituiu e tive a ventura de poder sempre contar com o seu estímulo intelectual demonstrando os seus invulgares conhecimentos matemáticos e com a sua orientação competente, empenhada e afetuosa, capaz de superar qualquer ameaça de desânimo...”

Na memória de todas as pessoas que conheceram Sousa Ramos permanecerá a natureza nobre do seu carácter humano feito com extrema simplicidade e modéstia, bem como a sua enorme e sistemática dedicação aos seus alunos, dia após dia, ou a lembrança do seu delicioso entusiasmo compartilhando a sua vasta cultura tanto em matemática quanto em física.

Além disso, José de Sousa Ramos deixa uma vasta e notável herança escrita, muito mais rica do que poderia sugerir a mera contagem de duzentos artigos em revistas internacionais com arbitragem. A maior parte dos seus artigos foram no campo de sistemas dinâmicos discretos e a partir de seu trabalho podemos ter uma noção do alcance da sua vasta cultura científica: equações funcionais, álgebras C^* , funções zeta, grupos hiperbólicos e muitos outros e, em particular, na dinâmica simbólica.

A importância do trabalho científico de José Sousa Ramos residia na sua grande capacidade criativa e na sua autonomia mental, às vezes colocando-o muito à frente dos seus colegas com pensamentos originais e reflexões posteriores, que, precisamente por essa razão, resultaram em transgressões do pensamento e, naturalmente, ideias para depois lapidar. Nada surpreendente para quem está habituado à reflexão científica, e sabendo que a verdade científica nunca irrompe como uma poção destilada e é, no princípio, uma conquista feita pelo uso imperfeito de ferramentas que serão aperfeiçoadas no futuro.

Podemos dizer que José de Sousa Ramos criou uma verdadeira escola de matemática que é, em si, uma rara e notável conquista, como a alcançada há meio século pelo eminente matemático português José Sebastião e Silva cujos efeitos perduram até hoje. Quem conheceu José Sousa Ramos, sempre o viu rodeado de estudantes. Tomamos a liberdade de o nomear Mestre. Sim, ele era o Mestre; em todos os níveis, começando a nível científico e terminando a nível pessoal.

Talvez tenham sido os seus ensinamentos nomeadamente na capacidade de ousar que me levaram a propor que esta disciplina incorporasse um plano curricular dum programa de doutoramento em Matemática e a escolher esta disciplina, onde se cruzam várias áreas científicas, para ser alvo do relatório das provas de agregação.

Uma frase dum poema de António Machado traduz bem a sua forma de estar:

“Caminhante não há caminho – o caminho faz-se caminhando”.

1.1.2 - Christian Mira

Christian Mira foi pioneiro no estudo de sistemas dinâmicos (comportamento equações recorrentes não lineares, em linha com o trabalho de Henri Poincaré em França, George David Birkhoff nos Estados Unidos e a escola soviética de Aleksandr Andronov e Lev Pontryagin), levando a sua equipa a pesquisas fundamentais sobre este tema com Igor Gumowski da Universidade Laval de Quebec.

Tive o privilégio de ter sido orientada no meu trabalho, em França, pelo Prof. Christian Mira (1998/99) e de ter beneficiado do seu saber. Devo-lhe a permanente disponibilidade para a permuta de ideias e os frequentes conselhos que tão cordialmente me concedeu. Abriu-me as portas para entrar no chamado “grupo de Toulouse” (que na altura se chamava Groupe d’Études des Systèmes Non Linéaires et Applications no INSA) que permitiu o estudo dos sistemas dinâmicos com uma visão abrangente, geométrica, computacional, também aqui aberta a enveredar por caminhos ainda não muito caminhados.

1.1.3 - Ton Coolen

Ton Coolen é atualmente professor na Radboud University, foi professor no King’s College London e na University of Oxford, sendo os seus principais temas de investigação activos o desenvolvimento e aplicação de métodos de inferência estatística e de aprendizagem automática para a análise preditiva de dados de alta dimensão em medicina (com ênfase na investigação do cancro), o desenvolvimento de métodos matemáticos e estatísticos para a modelação e análise de sistemas complexos de muitas variáveis (com aplicações em física, ciências biomédicas, ciência da computa-

ção e economia) e análise matemática de redes e gráficos aleatórios. Também trabalha com redes neuronais, processos estocásticos, imunologia e ciência de dados.

Foi durante a minha dispensa sabática em 2010 no Kings College London, trabalhando com Ton Coolen, professor no departamento de Matemática, sobre a noção de entropia em redes, que amadureci a ideia de propor para o curso de doutoramento de Matemática na Universidade de Évora uma disciplina de redes Dinâmicas. Nas múltiplas, desafiantes e gostosas conversas de trabalho, em áreas desde a Física, Biologia, Economia, Informática, Música e, obviamente Matemática e nos encontros semanais do Disordered Systems Group, tomei consciência que o nosso programa de doutoramento precisava deste reforço em termos da sua estrutura curricular. Ao trabalhar numa equipa multidisciplinar liderada pelo Ton, num projeto relacionado com a investigação do cancro com médicos do Guy Hospital, foi para mim evidente que, nodos matemáticos, devíamos envolver-nos em temas como este e para isso era necessário existir um suporte teórico que permitisse uma linguagem de intercâmbio.

1.1.4 - Mason Porter

É atualmente professor no Department of Mathematics, University of California, Los Angeles (UCLA) (desde 2016) e foi professor de Nonlinear e Sistemas Complexos no Mathematical Institute, University of Oxford. Tem interesse em diversas áreas da matemática aplicada e está sempre em busca de novas áreas para experimentar. Alguns destes temas são Dinâmica Não Linear e Caos; Ondas não Lineares, Caos Quântico, Ciência de Redes, Análise de Redes Sociais, Biologia Matemática (incluindo redes biológicas) e Sincronização. O seu percurso, de matemático especializado em Sistemas Dinâmicos que, entretanto, começou a trabalhar em teoria de Redes, levou-me a propor uma colaboração que se concretizou.

Durante a minha sabática na UCLA, em 2017, onde trabalhei com Mason sobre Geometria

Hiperbólica e Teoria de Redes, um dos temas das muitas sessões de trabalho que se caracterizaram por interessantes conversas sobre diferentes temas, científicos, pedagógicos, jornalísticos, foi o que deveria ser melhorado num curso que Mason lecionou, quando estava em Oxford, de Redes Dinâmicas. A sua forma de interagir com os seus alunos de doutoramento de diferentes países, provocando verdadeiras discussões científicas, acesas, nos encontros semanais do grupo que liderava, foi muito pedagógico. Na verdade esta estadia ajudou a implementar a lecionação desta disciplina.

1.2 A era da produção em larga escala de dados!

Nas últimas décadas, a nossa capacidade de colecionar e armazenar vastas quantidade de dados aumentou dramaticamente. Isso inclui dados socioeconómicos, ligações sociais entre indivíduos ou redes de colaboração profissional; dados de tráfego em sistemas informáticos e de telecomunicações, redes de satélites, internet, redes elétricas, ligações ferroviárias, rodoviárias ou aéreas e redes de distribuição.

Recolhemos e processamos grandes quantidades de dados geológicos e meteorológicos, dados sobre o nível do mar, poluição do ar e da água, registos vulcânicos e sísmicos e tamanhos de camadas de gelo polar e glacial. Finalmente, vimos uma explosão nos últimos anos de dados biomédicos, como dados experimentais sobre bioquímica, processos e estruturas nos níveis celular, sub-celular e mesmo molecular, as topologias de sistemas compostos complexos de sinalização e processamento de informações, como o cérebro ou o sistema imunológico, dados epigenéticos, dados epidemiológicos e grande número de registos de pacientes com informações clínicas.

Alguns factos actuais: a internet tem 2400 milhões de utilizadores, 50000 milhões de páginas da Web indexadas e mais de 600 milhões de servidores Web e, a partir de 2022, aproximadamente 97 por cento das empresas estão a investir no poder crescente do *big data*. Na figura 1.1 podemos observar a Internet em exibição com 5 milhões de arestas



Figura 1.1: Internet com 5 milhões de arestas.

Deve haver uma razão para armazenar esse vasto volume de dados. Deveria ser, o desejo de entender o comportamento dinâmico do sistema complexo que gerou os dados, para prever com precisão razoável a sua evolução futura e a sua resposta a perturbações ou intervenções, ou para compreender a sua formação. Podemos querer melhorar e otimizar a eficiência de um sistema, projetar controles regulatórios eficazes ou, no caso da medicina, entender e tratar, ou melhor, prevenir as doenças.

Infelizmente, o que acontece atualmente no nosso Mundo faz-nos pensar que, nem sempre, esses são os objectivos que orientam as aplicações do progresso científico.

Quando estudamos um processo dinâmico, preocupamo-nos com o seu comportamento como função do tempo, espaço e seus parâmetros. Existem numerosos estudos que examinam quantas pessoas estão infectadas por um contágio biológico e se ele persiste de uma estação para outra.

Todos esses estudos têm algo em comum: a dinâmica está ocorrendo em um conjunto de entidades discretas (os nodos numa rede) que estão conectadas umas às outras através de arestas de alguma maneira não trivial. Isso leva à pergunta natural de como a conectividade não trivial subjacente afeta os processos dinâmicos. Esta é uma das questões mais importantes em ciência de redes.

É importante realçar que não só a estrutura da rede pode afetar a dinâmica dos processos

numa rede, mas também que os processos dinâmicos podem afetar a própria dinâmica da rede. Por exemplo, quando uma criança fica com gripe, ela pode não ir à escola por alguns dias, e essa mudança temporária na atividade humana afeta quais os contatos sociais que ocorrem, o que pode, por sua vez, afetar a dinâmica da propagação da doença.

Para sistemas pequenos e simples a tradução da observação em compreensão qualitativa e quantitativa do design e da função geralmente não é difícil.

No entanto, se recolhermos dados sobre sistemas complexos com milhões de variáveis que interagem de forma não linear, apenas tendo uma lista das suas partes, das suas conexões, as observações de seus padrões coletivos de comportamento já não são suficientes para entender como esses sistemas funcionam.

1.3 A matemática na vanguarda do conhecimento!

À medida que enfrentamos desafios globais como mudanças climáticas, pandemias e a necessidade de fontes de energia sustentáveis, a matemática continua a ser um pilar para encontrar soluções inovadoras. Com a explosão de dados e o advento da era digital, a matemática tornou-se ainda mais crucial. A matemática, na vanguarda do conhecimento, não é apenas uma ferramenta; é um motor do progresso humano, capaz de transformar a maneira como vemos e interagimos com o mundo. Mas para que continue a desempenhar um papel central no avanço do conhecimento humano, os matemáticos devem ser ativos, colaborativos e críticos, ajudando a moldar as soluções para os desafios mais urgentes da sociedade contemporânea. Transcrevo dois textos que enquadram estes desafios.

Consideremos nas afirmações de Jürgen Jost:

“.....Portanto, como a recolha de dados hoje em dia está tipicamente à frente da compreensão teórica, a matemática deveria enfrentar radicalmente a falta de teoria e olhar para o que existe,

os dados, e ver o que se pode fazer com eles. E aquilo a que pretendo apelar não são métodos ad-hoc para cada conjunto de dados concretos, mas sim uma análise abstrata dos desafios estruturais mais profundos. É claro que a matemática moderna está desenvolvida e sofisticada o suficiente para fornecer ferramentas apropriadas e úteis para basicamente qualquer conjunto de dados, mas esta, por si só, é uma perspectiva científica demasiado estreita. Para mim, como matemático, a matemática é mais do que análise de dados. Precisamos de um repensar conceptual... [Jos 15].

...Uma perspectiva tradicional consideraria a matemática apenas como uma ferramenta quando conjuntos de dados (possivelmente muito grandes) fossem analisados. Uma das teses deste ensaio é que esse próprio processo, a análise de dados, pode e já se tornou objeto de pesquisa matemática. Assim, a matemática não só serve como uma ferramenta poderosa, mas ao refletir esse papel, ganha uma nova perspectiva e se eleva a um nível superior de abstração. Então, o domínio de tal investigação matemática não é mais um campo específico, como a física, mas a própria técnica matemática. A matemática deixa então de ser, se é que alguma vez foi, uma mera ferramenta formal para a ciência, mas torna-se a ciência das ferramentas formais..., [Jos 16]

(Jurgen Jost, Diretor do Max Planck Institute for Mathematics in the Sciences in Leipzig desde 1996).

Peter Grindrod, Desmond J. Higham, Robert S. MacKay, Opportunities at the Mathematics/Future Cities Interface, November 2014.

*“... No preâmbulo de seu recente livro *The New Science of Cities* [3], Michael Batty, do Bartlett Centre for Advanced Spatial Analysis da University College London, discute três princípios centrais onde explica a sua perspectiva de “redes e fluxos” da ciência da cidade. O livro de Batty faz uso de conceitos como modelação baseada em agentes, teoria dos grafos, cadeias de Markov, problemas de decisão markovianos, otimização e auto-similaridade/fractais e é, portanto, um excelente ponto*



Figura 1.2: Em 2000, metade da população mundial vivia em cidades e as Nações Unidas projectam que atinja 60% em 2030 e que, para o ano 2050, 70% da população mundial seja urbana.

de partida para matemáticos que desejam entrar na área...

1.4 O século da complexidade !

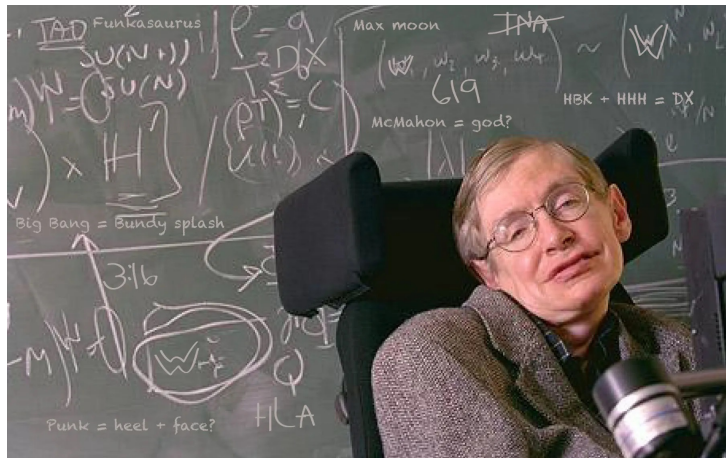


Figura 1.3: Sephen Hawking.

O século XXI tem sido amplamente referido como o "século da complexidade", uma afirmação que ganhou destaque nas palavras de Stephen Hawking. De acordo com Hawking, este será o

período em que a ciência e a tecnologia se concentrarão na compreensão dos sistemas complexos. Ele previu que, à medida que a ciência evolui, será necessário abordar fenômenos que envolvem interações complexas, como o comportamento das redes sociais, a dinâmica do clima e até mesmo as estruturas biológicas no nível molecular.

Essa visão de Hawking ecoa em diversos campos, como a teoria dos sistemas dinâmicos, a inteligência artificial e a modelação de redes complexas, onde a matemática desempenha um papel fundamental.

Além disso, este século tem visto o aumento exponencial de dados gerados por tecnologias emergentes. A compreensão de como esses dados interagem num nível macro, levando em conta as interações complexas, é uma das grandes fronteiras do conhecimento. Por outras palavras, a "ciência da complexidade" é uma chave para a inovação futura.

A percepção de que estamos inseridos e cercados por diversas redes e que estas possuem um papel central em diversos aspetos de nossas vidas, vem crescendo vertiginosamente durante a última década, tanto no meio académico quanto no público em geral.

1.5 A razão da existência da disciplina Redes Dinâmicas

Com um trabalho pioneiro na evolução de grafos aleatórios, a teoria dos grafos é frequentemente citada como a base matemática de ciência de redes. Apesar desta afirmação, as duas comunidades de investigação ainda estão em grande parte desligadas. É necessária uma interligação entre campos - preenchendo a lacuna entre grafos e redes - e como a ciência de redes pode beneficiar de tal influência. A ciência de redes, através de uma abordagem mais matemática, pode sugerir leis subjacentes de comportamento e prever redes complexas que descrevem fenômenos da natureza ainda não observados. Ao incorporar ferramentas de diversas áreas - como teoria dos grafos, álgebra

linear computacional, sistemas dinâmicos, otimização, física estatística, probabilidade e outras -, torna-se essencial para aplicações em quase qualquer campo imaginável. Assim, a inserção de disciplinas dedicadas ao estudo das redes nos currículos de física, estatística, ciência da computação, ciências sociais e organizacionais revela-se de suma importância, com destaque especial para os cursos de graduação e pós-graduação em Matemática. O objetivo é fomentar a coabitação harmoniosa entre a modelação matemática e o enquadramento teórico, dotando os alunos de conhecimentos e ferramentas que lhes permitam decifrar e compreender a literatura de investigação contemporânea.

Ensinar esta disciplina de redes dinâmicas tem sido muito desafiante. Os principais desafios incluíram encontrar o equilíbrio certo, especialmente dadas as diversas formações dos alunos, entre rigor matemático e modelos, métodos analíticos e computacionais e, numa fase seguinte, relacioná-los com conjuntos de dados reais.

Alguns problemas e miniprojetos do curso requerem a aplicação de recursos do mundo real, envolvendo grandes conjuntos de dados, o que dificulta a tarefa de torná-los igualmente acessíveis a estudantes com variados graus de competência computacional.

1.6 Estudo do panorama geral dos cursos existentes

Tendo em conta que se trata de um assunto tão abrangente e recente, a escolha da estrutura curricular é delicada. Neste sentido efectuei uma grande pesquisa nos diversos cursos que existem a nível internacional e nacional. Congressos como a NESTSCI, que reúne muitos dos membros da comunidade científica desta área, têm uma seção inteiramente dedicada ao Ensino em Ciência das Redes. Destaco um detalhado e sistemático estudo (por Sayama em 2016) que decorre destes encontros, o qual visou recolher e estudar, utilizando a ciência de redes, uma série de cursos de Ciências das Redes existentes, bem como as suas estruturas curriculares, que poderemos consultar nos links seguintes:

1. <http://barabasi.com/book/network-science>
2. <http://bingweb.binghamton.edu/sayama/SSIE641/>
3. <http://faculty.nps.edu/rgera/MA4404.html>
4. <http://hornacek.coa.edu/dave/Teaching/Networks.11/>
5. <http://mae.engr.ucdavis.edu/dsouza/mae298>
6. <http://networksatharvard.com/>
7. <http://ocw.mit.edu/courses/economics/14-15j-networks-fall-2009/>
8. <http://ocw.mit.edu/courses/media-arts-and-sciences/mas-961-networks-complexity-and-its-applications-spring-2011/>
9. http://perso.ens-lyon.fr/marton.karsai/Marton_Karsai_complexnet.html
10. <https://cns.ceu.edu/node/31544>
11. <https://cns.ceu.edu/node/31545>
12. <https://cns.ceu.edu/node/38501>
13. https://courses.cit.cornell.edu/info2040_2015fa/

14. <https://iu.instructure.com/courses/1491418/assignments/syllabus>
15. <https://sites.google.com/a/yale.edu/462-562-graphs-and-networks/>
16. <https://www0.maths.ox.ac.uk/courses/course/28833/synopsis>
17. <https://www.coursera.org/course/sna>
18. <https://www.sg.ethz.ch/media/medialibrary/2014/11/syllabus-cn15.pdf>
19. <http://tuvalu.santafe.edu/aaronc/courses/5352/>
20. <http://web.stanford.edu/class/cs224w/handouts.html>
21. <http://web.stanford.edu/jugander/mse334/>
22. http://www2.warwick.ac.uk/fac/cross_fac/complexity/study/msc_and_phd/co901/
23. <http://www.ait-budapest.com/structure-and-dynamics-of-complex-networks>
24. http://www.cabdyn.ox.ac.uk/Network%20Courses/SNA_Handbook%202013-14.pdf
25. <http://www.cc.gatech.edu/dovrolis/Courses/NetSci/>
26. <http://www.columbia.edu/itc/sociology/watts/w3233/>
27. <http://www.cse.unr.edu/mgunes/cs765/>
28. <http://www-personal.umich.edu/mejn/courses/2015/cscs535/index.html>
29. <http://www.stanford.edu/jacksonm/291syllabus.pdf>
30. <http://www.uvm.edu/pdodds/teaching/courses/2016-01UVM-303/>

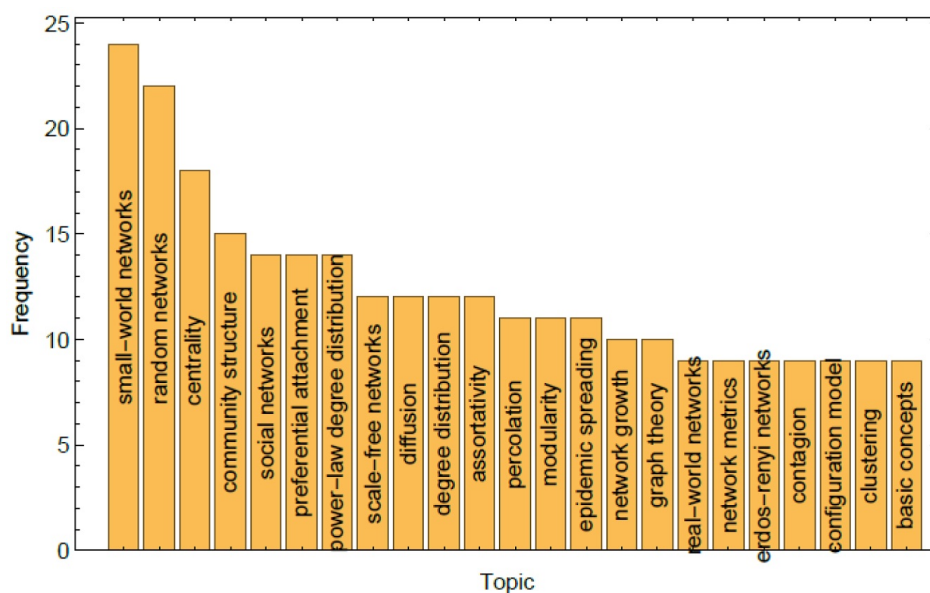


Figura 1.4: Frequências dos 20 principais tópicos que apareceram com maior frequência nos 30 programas/horários de cursos (os empates foram incluídos, de modo que um total de 23 tópicos aparecem neste gráfico).

As imagens na figura 1.4 ilustram alguns resultados da análise dos conteúdos programáticos destes cursos e na figura 1.5 visualizamos estas frequências como uma nuvem de palavras.



Figura 1.5: Visualização das frequências dos tópicos nos programas/horários dos 30 cursos como uma nuvem de palavras. Os tamanhos são definidos proporcionais com as raízes quadradas das frequências dos tópicos.

Pode observar-se que os capítulos: exemplos de redes (por exemplo, redes sociais), com alguns princípios básicos de estrutura de rede, introdução de redes aleatórias e redes de mundo pequeno

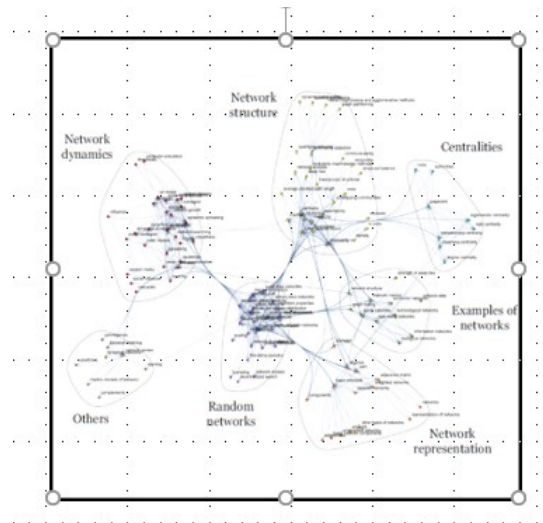


Figura 1.6: Estão identificados os sete clusters de tópicos na estrutura curricular dos cursos.

são comuns a todos os cursos e que qualquer um dos seguintes subtópicos pode ser considerado, dependendo do objectivo do curso:

- a. Redes sem escala e crescimento da rede
- b. Representação de redes
- c. Centralidades
- d. Outros tópicos sobre estrutura de rede
- e. Dinâmica de rede

No figura 1.6 são detectados sete clusters de tópicos na estrutura curricular dos cursos: (1) exemplos de redes, (2) representação de redes, (3) redes aleatórias, (4) estrutura de rede, (5) centralidades, (6) dinâmica de rede e (7) outros.

Capítulo 2

Descrição da disciplina Redes Dinâmicas

2.1 Objectivos e competências

Os estudantes de doutoramento devem tornar-se investigadores independentes e criativos, capazes de analisar criticamente e sintetizar ideias novas e complexas, e comunicarem com a comunidade científica e com a sociedade em geral, promovendo avanços científicos e tecnológicos e construindo uma visão crítica do conhecimento e da realidade. Como vimos, as redes são interessantes tanto matematicamente quanto computacionalmente, e estão difundidas na física, na biologia, na sociologia, na ciência da informação e numa infinidade de outros campos. O estudo de redes é uma das “estrelas em ascensão” do conhecimento científico, e as redes tornaram-se num dos assuntos mais importantes para os matemáticos aplicados estudarem. A maioria dos tópicos a serem considerados são áreas ativas de investigação moderna. Espero que, nesta disciplina, além de desenvolverem um sólido conhecimento e apreciação de algumas das ferramentas, conceitos e cálculos utilizados no estudo de redes, desenvolvam, também, a capacidade de ler e compreender artigos de investigação atuais na área. Neste sentido, e tendo em conta as experiências nas estadias em Londres, apresentei a proposta de inclusão, no programa de doutoramento em Matemática, da disciplina Redes Dinâmicas em 2011. Depois da avaliação pela A3ES deste programa doutoral, esta proposta foi

aceite. É publicada a estrutura curricular e o plano de estudos do curso. (Diário da República, 2.^a série, N.º 62, 30 de março de 2016).

2.2 Funcionamento da disciplina

A disciplina de Redes Dinâmicas foi leccionada pela primeira vez no ano lectivo de 2017/18. No relatório da acreditação, a identificação da disciplina é feita na tabela da figura 2.1.

Unidades Curriculares / Curricular Units	Área Científica / Scientific Area (1)	Duração / Duration (2)	Horas Trabalho / Working Hours (3)	Horas Contacto / Contact Hours (4)	ECTS	Observações / Observations (5)
Redes Dinâmicas	Matemática	Semestral	156	45	6	

Figura 2.1: Unidade curricular Redes Dinâmicas

No ano lectivo de 2017/18 funcionou em sistema E-learning com dois alunos inscritos: uma das alunas interrompeu o curso de doutoramento por questões pessoais e o outro aluno teve aproveitamento. No ano lectivo de 2018/19 funcionou em sistema E-learning com dois alunos inscritos, ambos os alunos tiveram aproveitamento. No ano lectivo de 2019/20 funcionou em sistema E-learning e presencial com dez alunos inscritos dos quais sete tiveram aproveitamento. No ano lectivo de 2020/21 funcionou em sistema E-learning e presencial com seis alunos inscritos, dos quais três tiveram aproveitamento. No ano lectivo 2024/25 está a funcionar com três alunas inscritas.

2.3 Programa Resumido

2.3.1 Grafos: generalidades e terminologia

- Grafo como estrutura discreta finita
- Tipos especiais de grafos .
- Conceitos de passeios, caminhos, circuitos, ciclos e comprimento.

- Grafos conexos, desconexos, fortemente e fracamente conexos, componentes conexas.
- Isomorfismo de grafos.
- Árvores: definição e propriedades.
- Representação de um grafo

2.3.2 Teoria espectral dos grafos.

- Revisão de alguns conceitos e resultados fundamentais de álgebra linear: espectro e raio espectral, propriedades, teorema Perron-Frobenius.
- Estudo do espectro de grafos isomorfos, completos, bipartidos completos e linha.
- Estudo de medidas de centralidade tendo como suporte a caracterização espectral das redes: número isoperimétrico e constante de Cheeger, centralidade do vector próprio, centralidade de um vértice por conectividade algébrica, centralidade de Katz e classificação de páginas (page-Rank).
- Exemplos de determinação destas medidas num caso simples.

2.3.3 Medidas de centralidade.

- Limitações e Percepções.
- Evolução e diversidade de aplicação das métricas de centralidade.
- Centralidade pontual.
- Medidas de centralidade no grafo.

2.3.4 Redes aleatórias

- “Rede aleatória” de Solomon e Rapoport, o chamado grafo aleatório de Poisson.
- O grafo aleatório de Erdős-Rényi (chamado grafo ER).
- Relação entre os modelos $G(n, m)$ e $G(n, p)$.
- Propriedades das redes aleatórias de Erdős-Rényi.
- Distribuição dos graus, diâmetro e comprimento médio de um caminho geodésico e suas implicações.
- Modelo de configuração.

2.3.5 Redes Small World

- Coeficiente Agrupamento (Clustering): Local, global.
- Experiência de Milgram e sucesso mediático.
- Modelo de redes Watts e Strogatz (WS).

Verificação deste fenómeno nalgumas redes.

2.3.6 Redes Sem-Escala

- Lei de potência.
- Redes "livre de escala".
- Modelo de "Anexação Preferencial" ("preferential attachment").
- Rede de Barabási-Albert.

- Comparação com redes aleatórias ER e com o modelo de Watts-Strogatz.

2.3.7 Partição de grafos e detecção de comunidades

- Noções de complexidade computacional.
- Partição de um grafo.
- Métodos espectrais, desigualdade de Cheeger.
- Método de Fiedler
- Detecção de comunidades.
- Algoritmo de Newman-Girvan.
- Optimização da modularidade através de passeios aleatórios.

2.3.8 Sistemas Dinâmicos em Redes e de Redes

- Distinção entre modelos para a dinâmica nas redes e a dinâmica das redes.
- Alguns sistemas dinâmicos estudados em redes.
- Percolação.
- Modelo Suscetível-Infectado-Recuperado (SIR).
- Contágios Sociais.
- Acoplamento de osciladores.
- Sincronização.

- Estudo da sincronizabilidade duma rede.
- Função zeta de um grafo.

2.4 Cronograma

Posteriormente, quando apresentarmos os sumários estendidos, serão detalhados os temas apresentados neste cronograma.

Módulos	Temas	Nº Aulas
1	Apresentação e motivação	1
2	Conceitos básicos	2
3	Teoria espectral	1
4	Medidas de centralidade	1
5	Redes Aleatórias e modelos de configuração	2
6	Redes “small world” e redes “Watt-Strogatz”	1
7	Redes "Sem escala"	1
8	Partição de redes e deteção de comunidades	2
9	Dinâmicas em Redes e de Redes	3
10	Apresentação dos trabalhos Finais	1

2.5 Metodologia e estratégias de ensino

Apresentação e discussão dos conceitos teóricos com rigor matemático e apelando à compreensão dos mesmos através do raciocínio lógico-dedutivo, da utilização das interpretações geométricas e das aplicações práticas desses conceitos. Resolução de exercícios com o objectivo de ilustrar a aplicação e manipulação dos conceitos necessários no estudo do comportamento de redes reais, com aplicação das metodologias seleccionadas e uso de software existente. Implementação de algoritmos seleccionados.

O contacto com investigadores e com as respectivas áreas de interesse de outras instituições irá complementar a formação académica dos estudantes e proporciona uma oportunidade de estabelecer contactos.

Quinzenalmente serão colocados na plataforma moodle uma lista de exercícios a resolver que incluem métodos computacionais.

Em cada módulo haverá um fórum de discussão, onde o aluno deverá colocar questões sobre a matéria e sobre os exercícios propostos.

2.6 Avaliação da Disciplina

A avaliação baseada em exames é particularmente inadequada para um curso de doutoramento e para uma disciplina em ciência de redes. Os problemas neste formato são artificialmente curtos e divergem substancialmente, tanto no tema quanto no tempo alocado, dos tipos de problemas que realmente devem pautar o início da atividade de investigação e do que se estuda na ciência de redes.

A preparação de trabalhos sobre a sua área de investigação e a sua apresentação promoverão a capacidade de comunicação oral do estudante, dando-lhes a oportunidade de explorar um tópico em profundidade e promover a pesquisas bibliográfica.

Para uma revisão de ferramentas de teoria dos grafos e conceitos iniciais de redes e para algumas metodologias de trabalho, foram indicados alguns exercícios que deveriam ser resolvidos e que constituíam uma primeira etapa no percurso desta disciplina.

Resumindo, a avaliação da unidade curricular é composta por dois momentos:

A - Lista de exercícios que serão resolvidos e posteriormente discutidos nas aulas,

B - Trabalho final, com um tema proposto pelo próprio aluno, com apresentação oral.

$$\text{Nota final} = 0.3 * A + 0.7 * B$$

2.6.1 Títulos dos trabalhos finais realizados pelos alunos

São elencados os títulos dos trabalhos finais realizados pelos alunos nos diversos anos.

- Estudo de medidas de centralidade na caracterização da cidade da Praia em Cabo Verde.
- Maximização Segmentada de Influência em Redes Sociais por Percolação Ótima.
- Estudo da estrutura da rede viária do Centro Histórico de Évora com teoria de grafos.
- Modelação baseada em Agentes e Redes Complexas
- Medidas Estacionárias em Grafos Infinitos.
- modelação baseada em Agentes e Redes Complexas.
- O Universo: a rede cósmica.
- Análise da Rede Facebook-686-potenciais *influencers*.
- Análise e Detecção de Comunidades
- Medidas de Centralidade e Análise Ferroviária: a Ferrovia para o comércio internacional e transporte no Laos/The Railway for International Trade and transportation in Laos
- Modeling Wildfire Behaviour using Complex Network Analysis while accounting for a Heterogeneous Orography and Forest Distribution / Modelação do comportamento dos incêndios florestais, usando análise de redes complexas, considerando uma orografia heterogênea na distribuição florestal.
- Modelo Epidemiológico SIR
- Rede de docentes de uma escola secundária.

- Teoria de Redes e Geometria Diferencial

Capítulo 3

Sumários estendidos

São apresentados os sumários estendidos de todos os módulos excepto no caso da primeira aula que é descrita na sua totalidade no Apêndice A deste documento.

3.1 Módulo 1 - Apresentação e Motivação

Como este módulo está apresentado no anexo A deste documento apenas transcrevemos os tópicos deste módulo.

3.1.1 Motivação

3.1.2 Programa

3.1.3 Funcionamento da disciplina

3.1.4 Avaliação

3.2 Módulo 2 - Conceitos básicos

Na aula anterior introduzimos informalmente a noção de rede e apresentaram-se vários exemplos. Para estudar redes, precisamos utilizar uma terminologia que nos permita ser precisos. Por exemplo, quando falamos sobre a distância entre dois nodos numa rede, o que realmente significa distância? Da mesma forma, é possível especificar quão bem é conectada uma rede?

Estas e outras afirmações podem ser formuladas com precisão usando a teoria dos grafos. Neste módulo o objectivo é fornecer um conjunto mínimo de notações, definições e conceitos usados na teoria de redes uniformizando linguagens. Algumas ilustrações e algumas proposições elementares serão oferecidas para cada conceito para uma melhor apreensão.

Com o objectivo de permitir uma melhor apreensão de conhecimentos iniciei a elaboração duma publicação que motivasse os alunos nesta desafiante área multidisciplinar e que harmonizasse notações, definições e conceitos básicos para servir de base ao estudo dos fundamentos da teoria dos grafos. Claro que conceitos e proposições mais complexos serão desenvolvidos em aulas que são dedicadas a assuntos substantivos específicos. Como material de apoio é fornecido um documento denominado “Notas introdutórias” onde estão organizadas definições e resultados essenciais.

3.2.1 Grafos: generalidades e terminologia

3.2.2 Grafo como estrutura discreta finita

Grafo como estrutura discreta finita: orientado / não orientado, ponderado / não ponderado, simples, regular, subgrafo induzido, ordem, tamanho, grau e sequência de graus.

3.2.3 Tipos especiais de grafos.

3.2.4 Conceitos de passeios, caminhos, circuitos, ciclos e comprimento.

3.2.5 Grafos e conexidade

Grafos conexos, desconexos, fortemente e fracamente conexos, componentes conexas.

3.2.6 Isomorfismo de grafos:

Estudo dos invariantes a menos de isomorfismo: ordem, tamanho, sequência de graus e subgrafos.

3.2.7 Árvores:

- Definição e propriedades.
- Árvore geradora.

3.2.8 Representação de um grafo

Esta representação pode ser especificada de diferentes formas equivalentes, por exemplo, gráfica, conjunto discreto: tabela dos pontos finais, tabela de incidência das arestas, lista dos vértices vizinhos, e matricial: matrizes de adjacência A , incidência M e laplaciana L .

3.3 Módulo 3 - Introdução à teoria espectral de grafos

Pretende-se relacionar propriedades algébricas do espectro de certas matrizes associadas a um determinado grafo e as propriedades estruturais presentes.

Neste trabalho vamos focar-nos no espectro da matriz de adjacência e no espectro da matriz laplaciana, ou seja, no fundo o objectivo é relacionar propriedades entre o grafo e o seu espectro. Como exemplo importante, estudamos o segundo menor valor próprio de uma matriz laplaciana que desempenha um papel relevante na teoria espectral, é chamado conectividade algébrica e sempre que esse valor for positivo então o grafo será conexo.

Uma aplicação interessante da conectividade algébrica é o estudo da sincronizabilidade nos vértices no acoplamento de sistemas dinâmicos, assunto que será abordado nas últimas aulas.

Revisão de alguns conceitos e resultados fundamentais de álgebra linear: espectro e raio espectral, propriedades, teorema Perron-Frobenius.

Estudo da centralidade do vector próprio e classificação de páginas.

Definição: Uma matriz não-negativa Q é irredutível se para qualquer i, j , existir k tal que $(Q^k)_{ij} > 0$ e diz-se primitiva se existir um k tal que para qualquer i, j , temos que $(Q^k)_{ij} > 0$.

Proposição. Q é irredutível se para qualquer par i, j , existe um caminho de j para i , $j \rightsquigarrow i$ e Q diz-se primitiva se $\exists k$ tal que para qualquer par i, j , existe um caminho $j \rightsquigarrow i$ de comprimento k .

Proposição: Uma matriz quadrada Q é irredutível se o gráfico orientado correspondente $G(A)$ for fortemente conexo (para quaisquer dois vértices P e Q , existir uma sequência de arestas (orientadas) que ligue P para Q).

Definição: Seja λ_i um valor próprio duma matriz A , de ordem $n \in \mathbb{N}$. A **multiplicidade algébrica** de λ_i , é a multiplicidade de λ_i como raiz do polinómio característico. A **multiplicidade geométrica** de λ_i , é a dimensão do respetivo subespaço próprio.

Revisão de alguns conceitos e resultados fundamentais de álgebra linear: espectro e raio espectral, matrizes semelhantes e propriedades, teoremas de Cayley-Hamilton e Perron-Frobenius.

Definição: Seja G um grafo não-orientado simples de ordem $n \in \mathbb{N}$, com o número de arestas $m \in \mathbb{N}$, o **espectro** de G , denotado por $esp(G)$, é a coleção das raízes do polinómio característico, isto é, os valores próprios da matriz de adjacência A tendo em conta as suas respectivas multiplicidades. Habitualmente, representamos o espectro por ordem decrescente, isto é, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, numa matriz linha.

Definição: Dado um grafo G simples não-orientado simples de ordem $n \in \mathbb{N}$, com o número de arestas $m \in \mathbb{N}$, o **raio espectral** denotado por $\rho(G)$ é um número real não negativo, tal que $\rho(G) = \max_{1 \leq i \leq n} |\lambda_i|$, sendo λ_i , $1 \leq i \leq n$ os valores próprios de G .

Definição: Seja G um grafo não-orientado simples de ordem $n \in \mathbb{N}$, com o número de arestas $m \in \mathbb{N}$, o **espectro da matriz laplaciana** L , denotado por $esp(L)$, é a coleção das raízes do polinómio característico, isto é, os valores próprios da matriz de adjacência L . Habitualmente, representamos o espectro por ordem decrescente, isto é, $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$, numa matriz linha.

Teorema de Perron-Frobenius: Seja A uma matriz real irredutível não negativa de ordem $n \in \mathbb{N}$. Então,

1. A tem um valor próprio real positivo igual ao seu raio espectral.
2. Esse valor próprio ρ_A tem multiplicidade algébrica e geométrica igual a um.

3. O único vector próprio x associado ao valor próprio ρ_A possui todas as entradas positivas.
4. ρ_A aumenta quando qualquer entrada de A aumenta.
5. Não existe outro valor próprio não negativo de A diferente de x .
6. Se, além disso, A é primitiva, então cada outro valor próprio λ de A satisfaz $|\lambda| < \rho_A$.

Casos particulares que relacionam a matriz de adjacência com a teoria espectral de grafos.

Apenas elencamos alguns resultados neste resumo.

Proposição: Seja G um grafo regular de grau r , então:

- (i) r é um valor próprio de G ;
- (ii) G é um grafo conexo se, e somente se, r tem multiplicidade 1;
- (iii) para todos os valores próprios de λ de G temos que $|\lambda| \leq r$.

Proposição: Sejam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ os valores próprios da matriz de adjacência A com n

vértices e m arestas, e seja T_h o número de caminhos de comprimento h em G , então:

- (i) A soma dos valores próprios é igual a zero, ou seja, $\sum_{i=1}^n \lambda_i = 0$;
- (ii) $T_h = \text{tr}(A^h) = \sum_{i=1}^n \lambda_i^h$;
- (iii) A soma dos quadrados dos valores próprios é duas vezes o número de arestas, ou seja,

$$T_2 = \sum_{i=1}^n \lambda_i^2 = 2m;$$
- (iv) Se G é um grafo regular de grau r então $T_2 = \sum_{i=1}^n \lambda_i^2 = rn$;
- (v) A soma dos cubos dos valores próprios é seis vezes o número de triângulos (t), ou seja

$$T_3 = \text{tr}(A^3) = 6t;$$

são estudadas outras proposições.

Casos particulares que relacionam a matriz laplaciana com a teoria espectral de grafos. Apenas elencamos algumas proposições neste resumo.

Proposição: Seja G um grafo regular de grau r com valores próprios da matriz de adjacência $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ e os valores próprios da matriz laplaciana $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$. Temos que

$$\mu_i = r - \lambda_{n-i+1}, \text{ para } i = 1, \dots, n.$$

Definição: Seja L a matriz laplaciana de um grafo G , então a **matriz laplaciana**, $\bar{L}$, do seu complementar $\bar{G}$ é definida por

$$\bar{L} = nI - J - L$$

onde J é a matriz quadrada de ordem n cujas entradas são todas iguais a 1 e I é a matriz identidade de ordem n .

Proposição: Seja G um grafo de ordem $n \in \mathbb{N}$ e v um vector próprio de L com valores próprios μ_i diferentes de 0. Então v é vector próprio de $\bar{L}$ com valores próprios $n - \mu_{n-i}$.

Teorema: Seja G um grafo não orientado simples de ordem $n \in \mathbb{N}$, com o número de arestas $m \in \mathbb{N}$. Sejam $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ os valores próprios da matriz laplaciana L . Então tem-se que:

1. $\mu_n = 0$;
2. G é conexo se e só se $\mu_{n-1} > 0$ (conectividade algébrica de G);
3. a multiplicidade algébrica de 0 é igual ao número de componentes conexas de G ;
4. $\sum_{i=1}^n \mu_i = 2m = \text{tr}(L)$.

Estudo do espectro de grafos isomorfos, completos, bipartidos completos e linha.

Estudo de medidas de centralidade tendo como suporte a caracterização espectral das redes: número isoperimétrico e constante de Cheeger, centralidade do vector próprio, centralidade de um vértice por conectividade algébrica, centralidade de Katz e classificação de páginas (page-Rank).

1. Número isoperimétrico

O número isoperimétrico de um gráfico é o número de arestas que devem ser removidas de um grafo para obter duas componentes conexas que são tão maior quanto é possível.

Definição: Seja G um grafo com n vértices ($\#V(G) = n$), $X \subset V(G)$ um subconjunto do conjunto dos vértices e $\delta X = \{(u, v) : u \in X, v \in V(G) \setminus X\}$ a fronteira de X . O **número isoperimétrico** de G é definido por:

$$i(G) = \min_{\#(X) \leq \frac{n}{2}} \frac{\#(\delta X)}{\#(X)}.$$

Este parâmetro está intimamente relacionado com a constante de Cheeger, uma importante constante estudada na geometria Riemanniana, que se pode definir em termos da teoria de grafos. O estudo da importância destes parâmetros na sincronização das redes será desenvolvido nos últimos módulos.

2. Centralidade do vector próprio

Os cientistas sociais usam a centralidade do vector próprio como forma de determinar líderes nas organizações. A centralidade do vector próprio tem por base os conceitos de valor próprio e vector próprio da matriz de adjacência do grafo. Esta medida diz-nos que uma maneira de interpretar a centralidade é pensar que a centralidade de um vértice está em função dos vértices vizinhos, ou seja, que um nodo importante tem vizinhos importantes

Definição: Seja G um grafo com n vértices, x um vértice de G , a_{jx} elementos da matriz de adjacência A e seja $\lambda_{\max}(a)$ o maior valor próprio (em módulo) da matriz de adjacência e $v = (v_1, \dots, v_n)^T$ o vector próprio associado. Definimos a **centralidade do vector próprio** por:

$$\sigma_E(x) = \sum_{j=1}^n a_{jx} \cdot v_j.$$

3. Centralidade por conectividade algébrica

Centralidade de um vértice por conectividade algébrica (menor valor próprio da matriz laplaciana não nulo): é utilizada para medir a importância de um vértice em relação à vulnerabilidade que ele oferece à rede caso tenha de ser dela retirado. Por outras palavras, deteta possíveis falhas que possam vir a comprometer a rede.

Definição: Seja G um grafo conexo e seja x um vértice de G , seja ainda $\alpha(G)$ a conectividade algébrica do grafo G , então $G \setminus x$ é o subgrafo induzido de G após retirado o vértice x e $\alpha(G \setminus x)$ é a conectividade algébrica de $G \setminus x$. Assim, a **centralidade do vértice x por conectividade algébrica** é definida por:

$$c_\alpha(x) = \alpha(G) - \alpha(G \setminus x).$$

4. Centralidade de Katz

Ao efetuar uma combinação linear do número total de caminhos de tamanho arbitrário de x para outros vértices, por exemplo, atribuindo um peso k^i aos caminhos de tamanho i , obtém-se a centralidade de Katz.

Definição: Seja G um grafo com n vértices, x um vértice de G , $\vec{1} = (1, 1, \dots, 1, 1)^T$, $\vec{e}_x = (0, \dots, 0, 1, 0, \dots, 0)^T$ e a matriz A^i cuja entrada a_{xy} representa o número de caminhos de comprimento i de x a y . Definimos a **centralidade de Katz** por:

$$\sigma_k(x) = \vec{1}^T \left(\sum_{i=1}^{\infty} k^i A^i \right) \vec{e}_x.$$

5. Classificação de páginas (Page Rank)

Page Rank é outra medida de centralidade bastante popular baseada em cadeias de Markov e passeios aleatórios. Este tópico é muito atual, a Google usa o Page-Rank para classificar os resultados da pesquisa, mede a importância de uma página contabilizando a quantidade e qualidade de links apontando para ela, a premissa do Page Rank é que um link de uma página para outra é um voto para a página de destino.

Exemplos de determinação destas medidas em casos simples.

3.4 Módulo 4 - Medidas de centralidade

Há uma ampla variedade de medidas de centralidade discutidas na literatura científica. Neste módulo, examinaremos algumas dessas medidas e proporemos orientações para futuras pesquisas

3.4.1 Limitações e Percepções

- O significado de centralidade não se limita apenas à forma como um vértice está conectado a outros vértices, mas também como pode controlar ou influenciar outros vértices na sua centralidade ou vulnerabilidade.

A centralidade de um vértice determina e influencia aspectos como conectividade, comunicabilidade e controlabilidade em uma rede. No entanto, a conectividade dos vértices nem sempre está alinhada com sua capacidade de gerar o tráfego (como a comunicabilidade), já que vértices com alta conectividade frequentemente enfrentam congestionamentos.

- Métricas de centralidade podem ser aplicadas em diversas disciplinas com finalidades diferentes. Além disso, existe um rico volume de métricas de centralidade disponíveis e utilizáveis para vários objectivos. Por exemplo, podemos querer investigar como equilibrar as cargas de tráfego, como definir limites entre nodos para tornar uma rede robusta contra falhas ou ataques, que tipos de ataques direccionados desenvolver, como identificar nodos vitais com base em vários critérios, ou o que é o nodo menos influente ou vulnerável numa determinada rede.

- Embora um grande volume de métricas de centralidade tenha sido desenvolvido até agora, apenas métricas de centralidade comuns foram usadas, como grau, intermediação, proximidade, coeficiente de agrupamento e pagerank, que foram desenvolvidas há várias décadas. Embora o grau seja uma métrica simples, outras métricas, como intermediação ou coeficiente de agrupamento, o

seu cálculo é muito demorado. Mesmo que tenha havido muitas centralidades métricas desenvolvidas na década de 2010, poucas delas têm sido usado nas aplicações de rede existentes enquanto as métricas desenvolvidas entre as décadas de 1970 e 1990 são as mais usadas na literatura.

3.4.2 Evolução e diversidade de aplicação das métricas de centralidade

A origem do desenvolvimento da centralidade métricas está ligada ao nascimento da teoria dos grafos, mas o estudo das métricas tem sido ministrado em muitas disciplinas, além da Matemática: química, antropologia, física, geografia, economia, psicologia, sociologia, biologia, gestão, informática, ciência política e psiquiatria. Identifica-se cerca de 60 medidas, no entanto apenas vamos considerar as que consideramos mais identificativas e importantes nas aulas seguintes.

A Figura 3.1 resume a evolução da centralidade em diversas disciplinas.

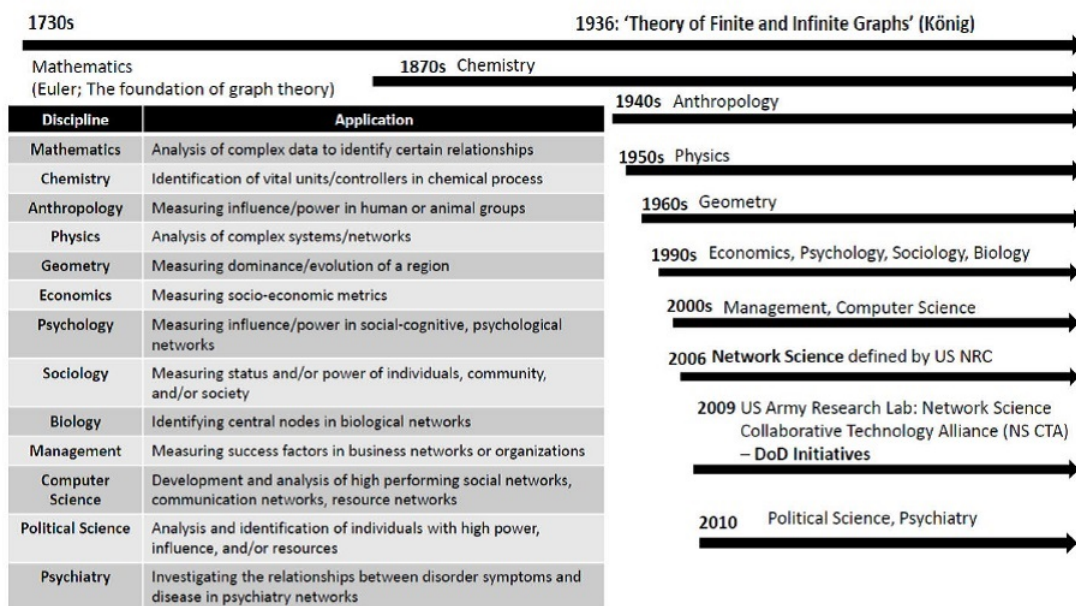


Figura 3.1: Evolução da centralidade em diversas disciplinas.

Medidas de centralidade agrupadas em duas classes - centralidade pontual medindo o papel do vértice na rede e centralidade do grafo medindo uma característica do grafo como um todo.

3.4.3 Centralidade pontual

3.4.4 Centralidade do grau

A medida de centralidade mais simples e conhecida é o grau do vértice ou o número de links ou arestas incidentes no vértice, ou seja, este parâmetro coincide com o próprio grau.

Definição: Seja o grafo G simples, conexo de ordem $n \in N$, v_i um vértice de $V(G)$. A centralidade do grau do vértice v_i é definida pela soma dos valores na linha i da matriz de adjacência, ou seja:

$$d(v_i) = \sum_j a_{ij}$$

A concepção mais simples e intuitiva no que diz respeito à centralidade de um vértice é o número de contatos diretos que ele possui. Uma pessoa que se encontra numa posição que permite o contato direto com muitos outros é vista pelos demais como um canal maior de informações, razão pela qual dizemos ser mais central.

3.4.5 Medidas baseadas no conceito de entropia

No contexto da termodinâmica, a entropia é uma medida de a desordem dos sistemas e no contexto da teoria da informação, a entropia mede a quantidade de informação ausente em um dado processo. Esses conceitos de entropia foram usados em redes, na caracterização de sistemas ou processos, utilizando-se várias variantes; vamos considerar duas variantes: a **entropia local** como a contribuição do vértice para a entropia da rede e a **entropia de mapeamento** para incorporar a consideração dos vizinhos do vértice;

Definição: Seja G um grafo simples, conexo de ordem $n \in N$, v um vértice de G , $d(v)$ o grau do vértice v e $N(v)$ o conjunto dos vértices adjacentes.

$$C_{entropia-local}(v) = - \sum_{u \in N(v)} d(u) \log d(u),$$

$$C_{entropia-map}(v) = -d(v) \sum_{u \in N(v)} \log d(u).$$

3.4.6 Índice h

Hirsch [92] introduziu o **índice** h com o objectivo de medir o impacto da produção científica de um investigador. Um investigador tem índice h se h for o maior inteiro l tal que o investigador tenha pelo menos l artigos, cada um com pelo menos l citações.

Como toda tentativa simplista de se classificar a produção científica de um investigador, por um único número, o índice h está longe de ser perfeito e enfrenta várias críticas

Mais tarde este índice h foi adaptado (chamando-o de **índice de lobby**) para descobrir vértices importantes nas redes: um vértice tem índice h se o vértice tiver pelo menos h vizinhos, cada um tendo pelo menos grau h , em que os restantes vizinhos só podem ter no máximo grau h . Estendendo este conceito, foi definido o operador $\mathcal{H}$ que, para qualquer vértice v , calculado nos graus do conjunto de seus vizinhos toma o número máximo h tal que h entradas tenham valor pelo menos h .

Definição: Seja G um grafo simples, conexo de ordem $n \in \mathbb{N}$, v um vértice de G e sejam $u_1, u_2, \dots, u_{d(v)} \in N(v)$.

$$C_{h-índice}(v) = \mathbf{h}(\mathbf{v}) = \mathcal{H}(d(u_1), d(u_2), \dots, d(u_{d(v)})).$$

Se o índice h de ordem zero do vértice v é o seu grau, ou seja, $h^{(0)}(v) = d(v)$, então o valor na equação anterior pode ser chamado de índice h de primeira ordem. Então k -ordem do índice h é definida como

$$h^{(k)}(v) = \mathcal{H}(h^{(k-1)}(u_1), h^{(k-1)}(u_2), \dots, h^{(k-1)}(u_{d(v)})).$$

Esta sequência sempre que a ordem aumenta, ou seja, $c_i = \lim_{k \rightarrow \infty} h^{(k)}(v)$.

3.4.7 Curvatura

O sucesso dos modelos hiperbólicos para redes, na reprodução de observações de redes reais, despertou muito interesse, nomeadamente no desafio de medir a geometria intrínseca de redes complexas.

Existem várias abordagens concorrentes para curvatura como (Eckmann e Moses [107]) da formulação de curvatura equivalente ao coeficiente de agrupamento local de Watts e Strogatz e é usada para revelar uma conexão entre alta curvatura e tópicos comuns na World Wide Web. Uma das abordagens mais utilizadas é derivada de uma curvatura gaussiana em gráficos planares, que foi generalizada para redes complexas.

Definição: *Seja G um grafo simples, conexo de ordem $n \in N$, v um vértice de G e s_v^k é o número de k -cliques incidentes em v . definimos:*

$$C_{\text{curv-Gauss}}(v) = \sum_{k \geq 0} (-1)^k \frac{s_v^{k+1}}{k+1}.$$

Uma versão truncada desta definição é usada para comparar um modelo de rede com dados. Uma terceira abordagem, de interesse recente, adapta uma noção de curvatura de Ricci às redes através da transferência de uma distribuição de massa de um vértice para outro e, portanto, pode ser definida numa aresta como $k(u, v) = W(m_u, m_v)$ onde $W(m_u, m_v)$ é o custo optimal de transporte de massa e a massa é normalmente um peso unitário distribuído proporcionalmente por um peso na aresta dos vizinhos aos vértices.

Definição: *Seja G um grafo simples, conexo de ordem $n \in N$, v um vértice de G . Definimos a curvatura num vértice v como uma soma ponderada da curvatura das arestas incidentes em v .*

$$C_{\text{curv-Ricci}}(v) = \frac{q}{d(v)} \sum_{u \in N(v)} k(u, v).$$

As seguintes medidas já foram estudadas no módulo da teoria espectral, apenas as elencamos como sistematização.

3.4.8 Centralidade do vector próprio

Esta medida é ocasionalmente chamada de centralidade de grau de Bonacich.

3.4.9 Centralidade de um vértice por conectividade algébrica.

3.4.10 Centralidade de Katz

Katz propôs uma nova medida considerando o número de conexões diretas a um vértice e os status dos vértices ligados a esse vértices.).

3.4.11 Classificação de páginas (page-Rank)

PageRank é uma variante moderna da centralidade de Katz que foi desenvolvida por Brin e Page, os fundadores do Google.

3.4.12 Coeficiente de aglomeramento

Esta medida será estudada com um ênfase especial no módulo seguinte dada a sua importância intrínseca e a sua importância no estudo dos modelos "Small World".

As medidas seguintes requerem a utilização, em geral, de toda a topologia da rede. Essas abordagens envolvem a medição de comprimentos de caminho entre vértices que estão separados (não adjacentes) na rede. Os cálculos de caminhos mais curtos geralmente não se adaptam bem ao tamanho da rede; por isso, estas médias são, geralmente, mais pesadas do ponto de vista computacional.

3.4.13 Centralidade de intermediação

Introdução [39] do conceito de intermediação parcial de um vértice numa dada rede para chegar a um valor que pudesse medir a centralidade deste vértice. Isto significa que este valor deveria ser capaz de expressar a influência que o indivíduo (representado pelo vértice) poderia exercer sobre os seus pares numa rede de comunicação do mundo real.

Definição: Seja G um grafo (conexo ou não) com n vértices e seja v_k um vértice de G . Considere um par de vértices v_i e v_j em G , tal que $i \neq j$, $i \neq k$ e $j \neq k$. A intermediação parcial de v_k com respeito a v_i e v_j é dada por:

$$b_{ij}(v_k) = \begin{cases} 0, & \text{se não existir caminho entre } v_i \text{ e } v_j; \\ \frac{g_{ij}(v_k)}{g_{ij}}, & \text{caso contrário,} \end{cases}$$

onde g_{ij} denota o número de geodésicas entre v_i e v_j e $g_{ij}(v_k)$ denota o número de geodésicas entre v_i e v_j que passam por v_k .

Para se determinar a centralidade de intermediação de um vértice v_k num grafo G com n vértices, basta somar todas as intermediações parciais de v_k em G .

Definição: Nas condições da definição anterior a centralidade de intermediação de um vértice v_k num grafo G com n vértices, denotada por $c_B(v_k)$ é dada por:

$$c_B(v_k) = \sum_{\substack{1 \leq i \leq j \leq n \\ i, j \neq k}} b_{ij}(v_k).$$

Este valor $c_B(v_k)$ indica o potencial que um ponto da rede, no caso v_k , tem para controlar o fluxo de informações entre os pares de vértices da rede. Quando a ordem n de G é um número muito grande, o procedimento de identificação e o de contagem de geodésicas, pode ser um trabalho

bastante moroso. Felizmente, existem resultados que permitem que estes cálculos sejam mais rápidos .

Teorema: (FREEMAN) *Seja G um grafo conexo ou não com $n \geq 3$ vértices. Seja v_k um vértice de G e $c_B(v_k)$ a centralidade de intermediação de v_k . Então:*

$$c_B(v_k) \leq \frac{n^2 - 3n + 2}{2}$$

e o limite é alcançado se e somente se G é o grafo estrela S_n e v_k é o vértice central da estrela.

3.4.14 Centralidade de Intermediação de Passeios Aleatórios

A centralidade de intermediação pode ser vista como uma medida da influência que um vértice tem sobre a propagação do fluxo de informação (ou de qualquer conteúdo) através da rede. Entretanto, em grande parte das redes, o fluxo pode ser governado por outras regras que consideram caminhos geodésicos e não geodésicos. Baseado nesta ideia mais realista de propagação da informação, foi introduzida uma medida de centralidade de intermediação baseada no fluxo de corrente em circuitos elétricos, para incorporar a contribuição de todos os caminhos (curtos e longos) com mais pesos atribuídos aos caminhos mais curtos.

Definição: *Seja G um grafo conexo com n vértices e seja v um vértice de G . A centralidade de intermediação de v_k em passeios aleatórios é dada por:*

$$c_{BA}(v_i) = \frac{2 \sum_{s < t} \tau_{st}(v_i)}{n(n-1)}$$

onde $\tau_{st}(v_i) = \frac{1}{2} \sum_j a_{ij} |p_{st}(v_i) - p_{st}(v_j)|$ para $v_i \neq v_s, v_y$, sabendo que a probabilidade que um passeio aleatório passe ao longo da aresta entre v_i e v_j é dado pela diferença absoluta entre as

probabilidades $|p_{st}(v_i) - p_{st}(v_j)|$ e a probabilidade de passar através do vértice v_i é a metade da soma das probabilidades nas arestas incidentes a ele:

3.4.15 Centralidade de proximidade

Bavelas [3] estava interessado em distinguir entre diferentes posições em redes de pequenos grupos. A mais simples e natural destas medidas, proposta por Sabidussi, em 1966, é baseada na soma das distâncias de um vértice em relação aos demais vértices do grafo.

Definição: Seja G um grafo conexo com n vértices e seja v_k um vértice de G . A centralidade de proximidade de v_k é dada pelo inverso da soma das distâncias de v_k a todos os outros vértices do grafo, ou seja,

$$C_C(v_k) = \frac{1}{\sum_{1 \leq j \leq n} d(v_j, v_k)}.$$

3.4.16 Centralidade de eficiência

Os problemas de determinação dum local de modo que minimize o tempo máximo de viagem entre o mesmo e todas as demais localizações é um assunto muito estudado, por exemplo, a instalação de um hospital, cujo objectivo é minimizar o tempo máximo de atendimento de uma ambulância a uma possível emergência. Neste sentido foi proposta, uma medida chamada centralidade eficiência, baseada no conceito de excentricidade de um vértice.

Definição: Seja G um grafo conexo com n vértices e seja v_k um vértice de G . A centralidade de eficiência de v_k é dada pelo inverso da excentricidade de v_k , isto é,

$$c_{efi}(v_k) = \frac{1}{ecc(v_k)},$$

onde $ecc(v_k) = \max_{v_i \in N(v_k)} d(v_i, v_k)$.

3.5 Medidas de centralidade no grafo

Agora olhamos para a centralidade de um determinado grafo, que caracteriza a centralidade de todos os vértices numa rede total e não apenas nos pontos (ou vértices). Portanto, a centralidade do grafo pode ser usada como um indicador para representar como os vértices de uma rede são conectados como um todo.

3.5.1 Número isoperimétrico

O número isoperimétrico de um grafo é o número de arestas que devem ser removidas de um grafo para obter duas componentes conexas que são tão maior quanto é possível.

3.5.2 Assortatividade de grau

A assortatividade de uma rede foi definida (Newman) como uma medida no grafo para representar até que ponto os vértices estão associados a outros vértices em termos de características estruturais da rede, como grau, intermediação, peso do vértice, bem como características do vértice, de acordo com algum contexto, como idioma ou cultura.

A medida de assortatividade mede a propensão de vértices se conectarem a outros vértices considerando características em comum, ou seja, a medida analisa se os vértices dessa conexão se comportam de forma similar ou não, verificando os graus dos vértices. O coeficiente de assortatividade r , é o coeficiente de correlação de Pearson entre pares de nodos ligados, e pode assumir valores entre $[-1, 1]$, de forma que valores positivos representam maior probabilidade dos vértices se comportarem de forma similar e, valores negativos maior probabilidade de que os vértices estejam ligados de forma distinta

Definição: Seja G um grafo com n vértices, m arestas e $A = [a_{ij}]$ a matriz de adjacência. O

coeficiente de assortatividade r é definido por

$$r = \frac{m^{-1} \sum_{j>i} k_i k_j a_{ij} - \left[m^{-1} \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij} \right]^2}{m^{-1} \sum_{j>i} \frac{1}{2} (k_i^2 + k_j^2) a_{ij} - \left[m^{-1} \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij} \right]^2}.$$

3.5.3 Curvatura do Grafo

Uma hipótese para explicar o fenómeno observado em muitas redes grandes com congestionamento de tráfego que ocorre num conjunto central de vértices da rede é que a rede como um todo é "*tem curvatura negativa*". As evidências que apoiam esta hipótese incluem a facilidade de incorporar redes no espaço hiperbólico ou derivar várias propriedades usando modelos de redes hiperbólicas, assunto muito estudado actualmente.

Se a rede tiver curvatura negativa, então os caminhos de roteamento influenciados pela seleção do caminho mais curto serão de certa forma forçados a atravessar esse núcleo, levando ao congestionamento. As centralidades pontuais são úteis para identificar potencialmente esse conjunto principal, mas não medem a curvatura da rede do grafo como um todo.

Foi desenvolvida uma medida de curvatura, adaptando aos grafos a "*condição de triângulo δ -fino*" ([167]) que define a curvatura negativa.

Para qualquer triplo de nodos i, j, k , define-se a função de distância de qualquer outro vértice h ao triângulo com estes vértices por

$$D(h; i, j, k) = \max \{d(h; i, j), d(h; i, k), d(h; j, k)\}$$

onde $d(h; u, v)$ é a distância mínima do vértice h à geodésica entre u e v .

Definição: Nestas condições, define-se a curvatura de uma rede em relação ao triplo, $\delta_{i,j,k}$,

pode ser definida como

$$\delta_{i,j,k} = \min_h D(h; i, j, k).$$

- Estudo do comportamento das medidas apresentadas considerando grafos orientados.
- Aplicação de muitas destas medidas a redes construídas pelos alunos e a redes reais.

3.6 Módulo 5 - Redes aleatórias

“Um matemático é uma máquina de transformar café em teoremas. (Paul Erdos)

A necessidade de “*modelos nulos*”, para saber quais valores esperaríamos encontrar por padrão ou normalmente.

Se observarmos que um grafo tem uma distribuição de graus de Poisson, deveríamos ficar entusiasmados?

Se descobrirmos que o número de triângulos num grafo de N nodos é igual a $N = 5$, este é um número grande ou pequeno?

Quais as características dos espectros de matriz de adjacência que são comuns à maioria das redes e quais são informativos e especiais?

Podemos definir valores “típicos” como aqueles que encontraríamos num “*modelo nulo*”, que definimos como um grafo aleatório que é semelhante à rede em questão.

3.6.1 “Rede aleatória” de Solomonoff e Rapoport

Referência à primeira tentativa séria de construir um modelo para redes grandes que foi a “rede aleatória” de Solomonoff e Rapoport, o chamado grafo aleatório de Poisson.

3.6.2 O grafo aleatório de Erdős-Rényi

O grafo aleatório de Erdős-Rényi chamado grafo **ER**

Definição: Um grafo **ER**, $G(n, p)$, é definido por n , número de vértices e p , probabilidade de que cada aresta simples (i, j) exista. Ou seja, as arestas são independentes e variáveis aleatórias identicamente distribuídas, e o tamanho da rede e a probabilidade de aresta especificam totalmente o modelo.

Matematicamente, $G(n, p)$ define uma família, ou uma distribuição paramétrica sobre todos os grafos $Pr(G|p)$. Escolhendo p parametrizamos esta distribuição, e desenhamos qualquer grafo G com probabilidade $Pr(G|p)$. Este facto significa que quando descrevemos as propriedades seguintes de $G(n, p)$ normalmente descrevemos propriedades médias, e instantes individuais tenderão a flutuar em torno desses valores.

Também podemos definir o modelo $G(n, m)$, intimamente relacionado com $G(n, p)$.

Definição: O modelo de grafo aleatório $\mathbf{G}(\mathbf{n}, \mathbf{m})$, com n vértices, m arestas é obtido com a escolha de m pares de vértices uniforme e aleatoriamente entre os $n(n - 1)/2$ pares possíveis.

3.6.3 Relação entre estes modelos $G(n, m)$ e $G(n, p)$

3.6.4 Propriedades das redes aleatórias de Erdős-Rényi:

- Tradicionalmente um grafo simples, mas também facilmente **generalizado** para arestas **direccionadas**, com ou sem laços.
- A conectividade é “**homogénea**” e é inteiramente controlada pelo parâmetro p de “densidade” da aresta; as arestas são independentes e distribuídas de forma idêntica.
- A **distribuição de graus** $Pr(k)$ é uma distribuição de Poisson com média $c = p(n - 1)$.
- O **diâmetro** e a **distância geodésica média** são $\mathbf{O}(\log \mathbf{n})$, existem *caminhos curtos* entre a maioria dos pares de vértices, tornando os grafos ER “parecidos com os grafos "Small World", (estudados numa próxima aula).
- O **coeficiente de agrupamento**

- A **maior componente conexa** é proporcional a uma rede de tamanho $O(n)$ quando $c > 1$, e é extremamente pequena, contendo $O(1)$ vértices, quando $c < 1$ (existe transição de fase em $c = 1$).

3.6.5 Gerando uma rede $G(n, p)$

6. **Calcular o grau médio** de duas formas:

- Sabendo que para qualquer vértice determinado i , existem $n - 1$ outros vértices j possíveis, e cada um desses pares i, j está conectado com probabilidade p , tem-se:

$$\langle k \rangle = c = \sum_{j \neq i} p = p(n - 1).$$

- Sabendo que existem $\binom{n}{2}$ pares de vértices, que cada um deles existe com probabilidade p e que, então, o número esperado de arestas é $\langle m \rangle = \sum_{j=1}^n \sum_{j>i} p = p \binom{n}{2} = p \left(\frac{n(n-1)}{2} \right)$, tem-se que:

$$\langle k \rangle = \frac{n}{2} p \left(\frac{n(n-1)}{2} \right) = p(n-1).$$

3.6.6 Distribuição dos graus

Como as arestas em $G(n, p)$ são variáveis aleatórias identicamente distribuídas, tem-se que a distribuição de graus, é

$$\Pr(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

3.6.7 Diâmetro e comprimento médio de um caminho geodésico e suas implicações

Determinação duma estimativa do diâmetro de um grafo ER.

$$(c - 1)^{l_{\max}} = O(\log n).$$

Da mesma forma a também distância geodésica média é $O(\log n)$.

3.6.8 Qual é a aparência dos gráficos $G(n, p)$.

No caso de $c < 1$ (ultra-esparso), as redes são compostas inteiramente de componentes pequenas ou muito pequenas, principalmente árvores perfeitas. Em $c = 1$, muitas dessas pequenas árvores começam a conectar-se, formando uma grande variedade de componentes maiores, a maioria das quais ainda são árvores. No entanto, para $c > 1$, vemos surgir uma grande componente conexa (na verdade, uma componente gigante).

3.6.9 Modelo de configuração

O modelo de configuração, define um grafo aleatório com uma sequência de graus específica. Estas redes são mais ricas na sua variedade estrutural do que os grafos aleatórios ER, porque têm estruturas de graus mais realistas - na verdade, podemos até definir $\vec{k}$ como sendo exatamente a sequência de graus de uma rede G do mundo real.

3.6.10 Propriedades dos grafos aleatórios do modelo de configuração

Estes grafos têm propriedades, que são semelhantes em muitos aspectos às propriedades dum gráfico ER.

3.6.11 Gerando uma rede de modelo de configuração

Ao contrário de um gráfico ER, gerar uma rede de modelo de configuração não é trivial. Essa complexidade adicional quase sempre vale a pena devido à sua utilidade prática.

3.7 Módulo 6 - Redes Small World

3.7.1 Coeficiente Agrupamento (Clustering): Local, global

O coeficiente de agrupamento local c_i mede a razão entre o número de arestas entre os vizinhos de um dado e o número máximo possível de arestas entre esses vizinhos. Duma forma simples podemos afirmar que se $c_i = 0$ nenhum dos vizinhos do vértice i está conectado entre si, se $c_i = 1$ então os vizinhos do vértice i formam um grafo completo, ou seja, todos estão ligados e se $c_i = 0,5$ há 50% de hipóteses dois vizinhos estarem ligados.

Definição: Denotando por k_i o grau do vértice v_i e por $N_i = \{v_j : e_{ij} \in E\}$ o conjunto dos vizinhos de v_i , então para um grafo, podemos definir, o **coeficiente de agrupamento** c_i do vértice v_i e a **média dos coeficiente de agrupamento** c do grafo por:

$$c_i = \frac{2|\{e_{ij}\}|}{k_i(k_i - 1)}, \text{ com } e_{ij} \in E; v_j, v_k \in N_i \text{ e } c = \frac{1}{N} \sum_{i=1}^N c_i.$$

Podemos definir o coeficiente de agrupamento, utilizando a matriz de adjacência.

Definição: O coeficiente de agrupamento $c_i(A)$ do vértice i com grau ≥ 2 num grafo com N vértices e com matriz de adjacência A é definido por:

$$c_i = \frac{\text{número de pares de vértices ligados entre vizinhos de } i}{k_i(k_i - 1)} = \frac{\sum_{j,k=1}^N (1 - \delta_{jk}) A_{ij} A_{jk} A_{ik}}{\sum_{j,k=1}^N (1 - \delta_{jk}) A_{ij} A_{ik}}$$

Como os produtos de entradas da matriz de adjacência de um grafo podem ser usados para

identificar caminhos, podemos contar o número de caminhos fechados de um determinado comprimento, usando a matriz de adjacência.

Definição: O número L_l de caminhos fechados de comprimento $l > 0$ num grafo com N vértices e com matriz de adjacência A é dado por

$$L_l = \sum_{i_1=1}^N \dots \sum_{i_l=1}^N \left(\prod_{k=1}^{l-1} A_{i_k i_{k+1}} \right) A_{i_l, i_1} = \sum_{i=1}^N (A^l)_{ii} = \text{Tr} (A^l).$$

Definição: Considere um grafo simples G e um vértice v_i . Um triângulo em v_i é um subgrafo completo de G com exatamente três vértices, incluindo v_i . Um triplo em v_i é um subgrafo de exatamente três vértices e duas arestas, onde v_i é incidente com as duas arestas.

Definição: O número de triângulos T_i envolvendo o vértice i num grafo simples com N vértices e com matriz de adjacência A é definido como $T_i = \frac{1}{2} \sum_{j,k=1}^N A_{ij} A_{jk} A_{ik} \in \mathbb{N}$.

T_i , que também pode ser escrito como $T_i = \frac{1}{2} (A^3)_{ii}$, conta o número de caminhos fechados de comprimento três, nos quais o vértice i participa.

Observe que em grafos simples tem-se $c_i = \frac{2T_i}{k_i(k_i - 1)}$.

Definição alternativa razoável para o coeficiente de agrupamento global

Definição alternativa razoável para o coeficiente de agrupamento global, com base no número de triplos e triângulos num grafo, e suas relações

Generalizações do coeficiente de agrupamento:

grafos orientados, grafos pesados, bipartidos e modelos de configuração

Coeficiente de agrupamento nas redes aleatórias ER

Propriedade: Sabendo o grau médio $\langle k \rangle$ numa rede aleatória ER, $G(n, p)$, tem-se que coeficiente de agrupamento de uma rede aleatória ER $G(n, p)$ tem o valor constante p

$$c_i = \frac{\langle k \rangle}{n} = p$$

3.7.2 Estudo das implicações sobre a transitividade nestas redes

Motivação: experiência de Milgram e sucesso mediático

Várias experiências sociais sugerem que a maioria das pessoas no mundo está conectada apenas através de um pequeno número de links de "amigos". Este fenómeno tem sido chamado *Seis Graus de Separação* (a partir de uma ideia de uma história de 1929 de um escritor húngaro)

Mas esta ideia teve grande repercussão depois da chamada experiência de Milgram. Em 1967, Stanley Milgram, na época professor de psicologia social em Harvard, estava interessado em saber qual era a probabilidade de duas pessoas selecionadas aleatoriamente se conhecerem. Isso eventualmente levou à questão de até que ponto quaisquer duas pessoas estavam separadas uma da outra. A distância foi expressa em termos de "A conhece B, quem conhece C, quem conhece D...", e assim por diante. Por outras palavras, a separação era determinada pela cadeia de conhecidos através da qual uma pessoa acabaria por encontrar a outra.

Afirmações de Milgram

Cada um de nodos pertence a uma estrutura de "pequeno mundo".

Não é verdade, no entanto, que cada um dos nossos conhecidos constitua uma base igualmente importante de contacto com o mundo maior ".

São seis graus poucos ou muitos?

Se duas pessoas estão separadas, elas estão muito distantes?

Assim, quando falamos de intermediários, falamos de uma enorme distância psicológica entre os pontos inicial e o alvo, uma distância que parece pequena porque consideramos 5 uma pequena quantidade “tratável”. Não devemos pensar nos dois pontos como sendo separadas por 5 pessoas, mas sim separadas por “5 círculos de vizinhanças”.

Existência de outros algoritmos

Número de Erdos, número de Kevin Bacon

3.7.3 Exemplos de redes: Facebook

3.7.4 Modelo de redes Watts e Strogatz (WS), (Nature 1998)

As redes de mundo pequeno, de acordo com Watts e Strogatz, são uma classe de redes que são “altamente agrupadas, como redes regulares, mas possuem comprimentos de caminho pequenos, como grafos aleatórios.”

Construção do modelo de "Small-World" de Watts-Strogatz-Newmann

O que nos leva ao modelo Watts-Strogatz onde são avaliadas duas quantidades:

- a. A separação típica entre dois vértices, calculado como o comprimento dos caminhos mais curtos entre todos os pares de vértices;
- b. O coeficiente de agrupamento C .

Rede regular	Rede "small-world"	Rede aleatória
$p = 0$	$0 < p < 1$	$p = 1$
distância média elevada	distância média pequena	distância média pequena
coef. de agrupamento elevado	coef. de agrupamento baixo	coef. de agrupamento baixo

O modelo original de Watts-Strogatz

Considere um conjunto de n vértices $\{v_1, \dots, v_n\}$ e um número par k . Para garantir que o grafo terá poucas arestas, escolha $n \gg k \gg \log(n) \gg 1$.

1º) passo: Ordene os n vértices numa circunferência e, para cada vértice, ligue os primeiros $k/2$ vértices vizinhos no sentido horário e depois os $k/2$ vértices vizinhos no sentido anti-horário.

2º) passo: Com uma probabilidade p , troque a aresta (u, v) por uma aresta (u, w) , em que w é um vértice diferente de u e (u, w) é uma aresta ainda não existente do grafo.

Vamos referir-nos ao grafo aleatório de Watts-Strogatz por $WS(n, k, p)$.

Na verdade este modelo nunca se tornava num grafo aleatório verdadeiro, mesmo quando todas as arestas são reconectadas, uma vez que cada vértice ainda está anexado a metade de suas arestas originais.

2º) O modelo de Newman-Watts:

Começando com um modelo circular de n vértices em que cada vértice tem grau c , passamos por cada uma das arestas de cada vez e arestas extras, geralmente chamados de atalhos, são adicionados entre pares de vértices escolhidos aleatoriamente, mas nenhuma aresta é removida da rede subjacente.

O parâmetro p no modelo de "mundo pequeno" controla a transição entre o modelo circular ($p = 0$) e o grafo aleatório ($p = 1$), garantindo um intervalo entre esses valores, onde coexistem valor alto do coeficiente de clustering e um valor pequeno do comprimento médio dos caminhos, propriedades aparentemente contraditórias.

Resultado importante:

Para um valor pequeno de p , o coeficiente de agrupamento da rede ainda é quase o mesmo que o da rede regular original, e o comprimento médio do caminho diminui rapidamente aproximando-se da situação nas redes aleatórias. Então existe um amplo intervalo de p onde o comprimento médio do caminho é quase tão pequeno quanto o das redes aleatórias, mas o coeficiente de agrupamento é muito maior que o de uma rede aleatória.

3.7.5 Propriedades de rede de mundo pequeno

Considere-se um conjunto de n vértices $\{v_1, \dots, v_n\}$ e um número par k . Para garantir que o grafo terá poucas arestas, escolha $n \gg k \gg \log(n) \gg 1$.

Distribuição de graus:

1º) O modelo original de Watts-Strogatz:

$$\mathbb{P}j = \sum_{n=0}^{\min(j-k, k)} \binom{k}{n} (1-p)^n p^{k-n} \frac{(pk)^{j-k-n}}{(j-k-n)!} e^{-pk}.$$

2º) O modelo de Newman-Watts:

$$\mathbb{P}j = \binom{n}{j-2k} \binom{2kj}{n}^{j-2k} \left(1 - \frac{2kp}{n}\right)^{n-j+2k}.$$

Coeficiente de aglomeração:

1º) O modelo original de Watts-Strogatz:

$$c = \frac{3(k-1)}{2(2k-1)} (1-p)^3.$$

2º) O modelo de Newman-Strogatz:

$$c = \frac{3(k-1)}{2(2k-1)} (1-p)^3.$$

Exemplos de redes:

Uma surpreendente diversidade de redes são de "mundo pequeno" no sentido SW:

- Redes neuronais.
- Muitas redes sociais (ou seja, Facebook, IMDB, etc).
- Algumas redes estruturais, como redes elétricas e redes de estradas.
- Notavelmente, a Internet não é um mundo pequeno no sentido dos WS porque tem um pequeno

coeficiente de agrupamento.

3.7.6 Verificação deste fenómeno nalgumas redes

3.8 Módulo 7 - Redes sem-Escala

"Ricos cada vez mais ricos"

3.8.1 Motivação: redes internet

Apesar do modelo aleatório $G(n, p)$ ser o modelo mais popular e mais estudado para representar estrutura de redes, não captura aspectos fundamentais da estrutura presente em muitas redes reais como por exemplo a Internet.

O modelo de rede aleatória difere das redes reais em duas características importantes:

(A) Crescimento

As redes reais são o resultado de um processo de crescimento que continuamente aumenta o número de vértices. Em contraste, o modelo de rede aleatória assume que o número de vértices é fixo.

(B) Anexo Preferencial

Em redes reais, novos vértices tendem a ligar-se aos vértices mais ligados. Em contraste, os vértices nas redes aleatórias escolhem aleatoriamente os seus parceiros de interação.

3.8.2 Lei de potência

Apesar de se aplicar de forma mais geral, estamos interessados em distribuições de probabilidade que possuem este fenómeno, ou seja, em distribuições de lei de potência.

Definição: Uma distribuição é dita lei de potência quando sua função de probabilidade possui a seguinte forma:

$$p_X(x) \sim cx^{-\alpha} \text{ para } x \text{ suficientemente grande}$$

onde $p_X(x)$ é a probabilidade da variável aleatória discreta X assumir o valor x , c é a constante de normalização e α é uma constante e parâmetro da distribuição, com $\alpha > 1$.

Características importantes para o nosso estudo:

- (i) a distribuição de lei de potência decresce muito mais devagar do que qualquer outra distribuição, diz-se que uma distribuição de lei de potência possui “*cauda pesada*”.
- (ii) o seu gráfico traçado em escala $\log - \log$ é uma reta.

Exemplos da distribuição de Poisson e função zeta de Riemann

3.8.3 Função de distribuição *livre de escala*

Uma função distribuição de lei de potência também recebe o nome de *livre de escala*. Intuitivamente, a proporcionalidade da distribuição mantém-se em qualquer escala para os valores que ela pode assumir, sendo assim, livre de escala.

Definição: *seja b um valor inteiro positivo qualquer. Considere a razão entre um valor x qualquer para a distribuição e um valor b vezes maior, ou seja, bx . Temos então:*

$$\frac{p_X(x)}{p_X(bx)} \sim \frac{cx^{-\alpha}}{c(bx)^{-\alpha}} = b^\alpha \text{ para } x \text{ suficientemente grande}$$

3.8.4 Redes *livre de escala*

Definição: Uma rede é dita livre de escala se sua distribuição de grau segue uma lei de potência.

3.8.5 Modelo de *Anexação Preferencial (preferential attachment)*

No final da década de 90, Albert-László Barabási e Réka Albert propõem um modelo matemático para construção da Web tentando mostrar que esta característica, livre de escala, seria consequência do processo de formação da Web. Mas como podemos resumir este processo de construção da Web em que páginas novas que entram na Web tendem a criar hiperlinks para páginas mais

populares na Web e como definir popularidade? Barabási e Albert batizaram este mecanismo de *preferential attachment* e propuseram o seguinte modelo.

Mecanismo de *Anexação Preferencial* ("preferential attachment") de Barabási-Albert

1º) Considere o tempo discreto $t = 0, 1, \dots$. No tempo $t = 0$ temos a rede da Web inicial.

2º) A cada instante de tempo $t > 0$ um novo vértice (página) é adicionado à rede da Web. Este vértice "traz" com ele m arestas (hiperlinks) que ligam o próprio vértice a outro vértice da rede. A escolha deste vértice é feita de forma aleatória.

Seja u um vértice presente na rede no instante t e seja $d_u(t)$ o seu grau neste instante. A probabilidade de u ter sido escolhido é dada por:

$$p_u(t) = \frac{d_u(t)}{\sum_{v \in V(t)} d_v(t)}$$

onde $V(t)$ é o conjunto de vértices que existem na rede no instante t , (**Anexação preferencial**).

Desta forma, a probabilidade de u ser uma ponta da aresta de uma nova página depende de t .

3º) O vértice v_i que "entra" na rede no instante i , passa a fazer parte do conjunto de vértices nos instantes seguintes, ou seja, $v_i \in V_j$ para todo $j > i$.

3.8.6 Rede de Barabási-Albert

Propriedades

- **Propriedade:** Seja por $n(t)$ o número de vértices e $m(t)$ o número de arestas no instante t e n_0 e m_0 o número de vértices e o número de arestas na rede inicial, criada no instante $t = 0$. Considerando $m(t) = m_0 + mt$ e $n(t) = n_0 + t$ no instante t tem-se que o **grau médio da rede**, é dado por:

$$\bar{d}(t) = \frac{2m(t)}{n(t)} = \frac{2m_0 + 2mt}{n_0 + nt} \approx 2m \text{ para } t \text{ suficientemente grande.}$$

• **Propriedade:** Depois de um tempo t suficientemente grande a **distribuição de grau** da rede gerada pelo modelo BA será aproximadamente

$$P[d_u(t) = k] \approx \frac{2m^2}{k^3}.$$

Ou seja, a rede possui uma distribuição de grau de entrada que segue uma lei de potência com $\alpha = 3$.

• **Propriedade:** A **distância média**, calculada através do diâmetro da rede verifica:

$$\langle d \rangle \approx \frac{\log N}{\log \log N} \text{ para } N \geq 10^4.$$

• **Propriedade:** O **coeficiente de agrupamento** verifica:

$$c \approx \frac{(\log N)^2}{N}.$$

Na verdade, o mecanismo de anexação preferencial e suas consequências já haviam sido estudados por Simon na década de 60 no contexto de modelos econômicos, e por Price na década de 70 no contexto de redes de citação de artigos científicos [Simon 1955, de Solla Price 1965].

3.8.7 Exemplos de aplicação deste modelo

Exemplos de aplicação deste modelo e sua comparação com redes aleatórias ER e com o modelo de Watts-Strogatz.

3.9 Módulo 8 - Partição de grafos e detecção de comunidades

3.9.1 Motivação

Partição de grafos e detecção de comunidade são duas questões diferentes, mas bastante importantes e relacionadas na ciência de redes. Ambas estudam o problema de encontrar grupos (ou clusters) de vértices de forma que o número de arestas entre os grupos seja mínimo.

- O problema de partição de grafos ou agrupamento de grafos refere-se a uma classe geral de problemas que trata da seguinte tarefa: dado um grafo $G = (V, E)$, agrupar os vértices de um grafo em grupos ou clusters ou comunidades. O algoritmo de partição de grafos leva em consideração as propriedades do vértice e da aresta e , portanto, normalmente depende de algum tipo de métrica de “contagem de arestas” para otimizar.

Um exemplo típico da ciência da computação é a computação paralela onde vértices são tarefas de computação e arestas representam troca de dados entre tarefas. Existe um número fixo de CPUs e pretende-se equilibrar a carga de trabalho entre CPUs de tamanhos de grupo iguais.

Comunidades representam a organização de vértices em grupos, os quais podem ser identificados pela existência de muitas arestas ligando vértices de um mesmo grupo e poucas arestas ligando vértices de grupos diferentes. O processo de determinar as comunidades de uma rede complexa é chamado de detecção de comunidades. O objectivo da detecção de comunidade é encontrar as linhas "naturais" para a divisão de uma rede. Os tamanhos dos grupos podem variar amplamente de um grupo para outro.

O uso mais comum da detecção de comunidades é a compreensão dos dados da própria rede. Por exemplo, em redes sociais, detectar as comunidades sociais ou, nas redes metabólicas, detectar as unidades funcionais dentro da rede.

- Os algoritmos de partição de grafos preocupam-se apenas com o corte mínimo, não com a similaridade de vértices ou similaridade de conexão ou conexão densa. Além disso, nestes algoritmos, o número de partições deve ser predefinido.

Algoritmos de detecção de comunidades preocupam-se com a densidade, encontram a parte mais densa da rede e esse tipo de algoritmo não precisa predefinir o número de comunidades.

3.9.2 Noções de complexidade computacional

O pior caso de tempo de execução de um algoritmo para uma instância de problema de tamanho x é o número de etapas de computação necessárias para executar o algoritmo para a instância de tamanho x mais difícil possível. Da mesma forma, o pior caso de consumo de memória é o número de unidades de memória que o algoritmo precisará ocupar simultaneamente no pior caso possível para uma instância de tamanho x .

Na complexidade computacional, o interesse está em caracterizar como o tempo de execução e o consumo de memória crescem quando x cresce. Seja $f(x)$ uma função de x que determina o número de etapas de computação (ou alternativamente as unidades de memória) necessárias no pior caso, dado $x \in \mathbb{Z}^+$. A complexidade maior de um algoritmo é denotada por $O(g(x))$, onde $g(x)$ é uma função do tamanho da entrada x tal que $f(x)$ não cresce mais rápido que $g(x)$.

3.9.3 Partição de um grafo

O objectivo no particionamento de grafos é minimizar o número de arestas que ligam um subgrupo de vértices a outro, geralmente impondo limites ao número de grupos, bem como ao tamanho relativo dos grupos.

Uma questão importante é que o número e o tamanho dos grupos são fixos e às vezes os

tamanhos são fixos dentro de um intervalo, mas ainda são fixos. Quase todos os métodos podem ser descritos como combinação de métodos de:

- Melhoria Local (e multi-resolução).
- Métodos Espectrais.
- Métodos baseados no fluxo.

Melhoria Local (e multi-resolução)

Estes começam com uma partição e melhoram os cortes invertendo os vértices para frente e para trás. Informalmente, a ideia básica é que se houver algum tipo de geometria, digamos que o grafo que está sendo partido representa uma rede rodoviária, ou seja, que vive numa superfície bidimensional, então podemos “granular grosseiramente” sobre a geometria, para obter vértices e arestas efetivos e, em seguida, partir o grafo definido de maneira grosseira.

Algum contexto histórico: Kernighan-Lin nos anos 70; Fiduccia-Mattheyses nos anos 80, Chaco nos anos 90.

3.9.4 Métodos espectrais

Referem-se a uma classe de métodos que, na raiz, são um método de relaxação dum problema com um grau de complexidade grande e que envolve cálculos de vectores próprios.

Algum contexto histórico:

- Fiedler associou o segundo menor valor próprio do Laplaciano à conectividade do grafo e sugeriu dividir o grafo ao longo do vector próprio associado.

- Cheeger estabeleceu conexões com relações isoperimétricas em variedades contínuas, estabelecendo o que hoje é conhecido como a desigualdade de Cheeger.

Definição: A constante de Cheeger dum grafo G com o conjunto de vértices V é definida por

$$h(G) = \inf_X \frac{|\partial(X)|}{\min(\text{vol}(X), \text{vol}(\bar{X}))}$$

onde o ínfimo está sobre todos os subconjuntos X do conjunto de vértices e $|\partial(X)|$ é o número de arestas que unem os vértices de X com vértices de $V - X$.

Desigualdade de Cheeger: Sendo μ_1 , a desigualdade de Cheeger para um grafo G estabelece que

$$2h(G) \geq \mu_1 \geq \frac{h(G)^2}{2}.$$

- LLR considerou a geometria de grafos e aplicações algorítmicas, e interpretou LR como incorporação de G num espaço métrico.

- Chung concentrou-se no Laplaciano normalizado para gráficos de graus irregulares e na métrica associada de condutância.

- Shi e Malik usaram cortes normalizados para aplicações de visão computacional, que é essencialmente uma versão de condutância.

Ênfase nos métodos espectrais nomeadamente num elegante método devido a Fiedler que fornece boa aproximação para o problema da partição de grafos em duas partes.

3.9.5 Método de Fiedler

a) Seja A a $N \times N$ matriz de adjacência duma rede não orientada e não pesada. Queremos dividir os nodos em dois grupos C_1 e C_2 , tais que:

$$R = \frac{1}{2} \sum_{i \in C_1, j \in C_2} A_{ij}$$

seja tão pequeno quanto possível.

Encontrar a partição correcta é um problema de dificuldade NP.

b)

$$s_i = \begin{cases} +1 & , i \in C_1 \\ -1 & , i \in C_2. \end{cases}$$

$$\frac{1}{2}(1 - s_i s_j) = \begin{cases} 1 & , i, j \text{ estão em grupos diferentes} \\ 0 & , \text{ nos outros casos.} \end{cases}$$

As arestas entre os grupos são:

$$R = \frac{1}{4} \sum_{i, j} A_{ij} (1 - s_i s_j)$$

c) o que significa que:

$$R = \frac{1}{4} s^T L s = \frac{1}{4} \sum_{i, j} (k_i \delta_{ij} - A_{ij}) s_i s_j$$

d) O problema do particionamento de grafo está agora reduzido a

$$\min R = \frac{1}{4} s^T L s$$

e) Agora, temos um problema de optimização combinatória. Se relaxarmos o problema para que s_i possa tomar qualquer valor, temos o problema de optimização convencional.

f) Mostra-se que:

$$\frac{1}{4} s^T L s = \frac{1}{4} x^T L x = \frac{1}{4} x^T x = \lambda \frac{n_1 n_2}{N}.$$

g) Devemos escolher para λ um dos valores próprios de L . Não podemos escolher $\lambda_1 = 0$ porque o seu vector próprio é $\mathbb{I}$. Escolhemos, então, o segundo valor próprio mais pequeno λ_2 (a *conectividade algébrica* do grafo).

h) O seu vector próprio x é conhecido como o **vector de Fiedler**. E o valor de λ_2 é uma medida da qualidade das bi-partições da rede.

i) Voltamos agora atrás para $s_i = x_i + \frac{n_1 - n_2}{N}$, e devemos não "relaxar" os valores de s (tomar ± 1) para obter as colocações dos vértices em C_1 e C_2 .

j) nodos atribuímos os n_1 vértices, cujos $x_i + \frac{n_1 - n_2}{N}$ são maiores, a C_1 (i.e., $s_i = 1$) e o resto a C_2 (i.e., $s_i = -1$).

Esse método pode ser aplicado para obter um número arbitrário de partições e o seu sucesso é extamente por ser um relaxamento de um problema de otimização combinatória para o qual existem garantias de qualidade de aproximação no pior caso.

Métodos baseados no fluxo

Esta classe de métodos baseados no fluxo usa o procedimento de fluxo de múltiplos caminhos para revelar "engarrafamentos" no grafo. Intuitivamente, o fluxo deve ser "perpendicular" ao corte. A ideia é considerar um grande número de caminhos simultaneamente entre pares aleatórios de vértices e então escolher o corte com o maior número de arestas congestionadas.

3.9.6 Detecção de comunidades

Agora consideramos um problema que está relacionado, mas é diferente. Não podemos impor o número de grupos, devemos encontrá-lo assim como detectar os elementos desses grupos. Desta forma,

estamos observando a rede através duma lente para ver quão perto podemos chegar a uma estrutura de blocos diagonais numa matriz de adjacência. Procuramos encontrar a "melhor" atribuição de vértices a comunidades para alcançar a “melhor” estrutura de bloco-diagonal possível.

3.9.7 Tipos de Métodos

Existem centenas de métodos para tentar detectar comunidades algoritmicamente:

- **divisivo:** considera-se uma rede completa e divide-se recursivamente em pequenas grupos (até, por exemplo, cada vértice estar sózinho ou até algum critério de paragem).
- **aglomerado:** começa-se com vértices individuais e são combinados recursivamente até obter uma rede completa.
- **local:** obtém-se comunidades separadamente a partir de cada vértice (ou subconjunto de vértices) e, se for o que se pretende, tomando a média destes resultados.
- **clustering nas arestas:** atribuí-se as arestas às comunidades em vez dos vértices.

3.9.8 Algoritmo de Newman-Girvan

Este método clássico de deteção de comunidades usa uma versão da medida de centralidade de intermediação aplicada às arestas da rede.

Definição: Considere um par de vértices v_i e v_j e uma aresta l no grafo G . Denotamos por g_{ij} o número de geodésicas entre v_i e v_j e g_{ilj} o número de geodésicas que contém l . **A centralidade de intermediação da aresta l** é dada por:

$$c_B(l) = \sum_{i,j} \frac{g_{ilj}}{g_{ij}}$$

- **Algoritmo de Newman-Girvan**

- (i) Calcular c_B para todas as arestas da rede.
- (ii) Remover a aresta com maior c_B .
- (iii) Recalcular c_B para todas as arestas afetadas com esta remoção.
- (iv) Repetir a partir do passo (ii) até não sobraem mais arestas.

3.9.9 Modularidade

Em termos de detecção de comunidades, a medida mais conhecida é a modularidade. Tal medida, proposta por Newman e Girvan, está relacionada com a qualidade de uma determinada divisão da rede, é essencialmente a comparação entre o número de arestas de um determinado subgrafo da rede ou comunidade c e o número de arestas no modelo nulo, ou seja, um grafo aleatório de mesmo tamanho e sequência de graus da rede. Assim, um subgrafo é mais próximo de uma boa comunidade quando tem maior modularidade isto é, se o seu número de arestas internas exceder o número esperado de arestas internas que o mesmo subgrafo teria no modelo aleatório.

Definição: Seja G um grafo com n vértices e m arestas e $A = [a_{ij}]$ a matriz de adjacência. Seja k_i o grau de v_i e $\delta(c_i, c_j) = 1$ se v_i e v_j estiverem na mesma comunidade, e 0 no caso contrário. Denotemos por G^* uma divisão em comunidades do grafo. A **modularidade** de G^* é dada por:

$$Q(G^*) = \frac{1}{2m} \sum_i \sum_j \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j).$$

3.9.10 Optimização da modularidade através de passeios aleatórios

Vamos considerar um processo de passeio aleatório discreto no grafo G . A cada passo de tempo um caminhante está num vértice e move-se para um vértice escolhido aleatoriamente e uniformemente entre seus vizinhos.

Passeio aleatório

Consideremos $G = (V, E)$ um grafo conexo com n vértices e m arestas.

Começamos num vértice v_0 , se na t -ésima etapa estamos num vértice v_t , movemos o vizinho de v_t com probabilidade $\frac{1}{d(v_t)}$. A sequência de vértices aleatórios $(v_t : t = 0, 1, \dots)$ é uma cadeia de Markov.

Sendo A a matriz de adjacência de G , o vector contendo os graus dos vértices é $d = A1$, onde 1 é o vector $n \times 1$ de uns, D a matriz diagonal e M a matriz de transição e P o vector de probabilidades de dimensão $1 \times n$, a dinâmica do **passeio aleatório** é expressa pela equação:

$$P_{t+1} = P_t M.$$

Distribuição estacionária e estabilidade de Markov da partição

Supondo que o grafo não é bipartido esta dinâmica converge para uma única **distribuição estacionária**

$$\pi = \frac{d^T}{2m}$$

Cada partição do gráfico em c comunidades é codificada por uma matriz indicadora $n \times c$, H , com $H_{ij} \in \{0, 1\}$, onde 1 significa que o vértice i pertence à comunidade j . Sendo $\Pi = \text{diag}(\pi)$, a matriz de agrupamento do processo de difusão no tempo t é:

$$R_t(H) = H^T [\Pi M^t - \pi^T \pi] H$$

A $c \times c$ matriz $R(t)$, reflete a probabilidade do passeio aleatório permanecer dentro de cada bloco (elementos diagonais) e de existir transferência entre blocos (fora dos elementos diagonais) após um tempo t . Definimos a **estabilidade de Markov** da partição H como

$$r_t(H) = \min_{0 \leq s \leq t} \text{Tr}[R_s(H)] \approx \text{Tr}[R_t(H)].$$

Optimização da modularidade

A estabilidade de Markov $r_t(H)$ pode ser usada como uma função de maximização, para cada tempo t , no espaço de todas as partições possíveis do grafo

$$r_t = \max_H [r_t(H)]$$

Embora esta otimização seja NP -difícil, uma variedade de eficientes heurísticas de otimização para agrupamento de grafos podem ser usadas. A heurística do modelo de Potts proposta por Reichardt & Bornholdt [47] é recuperada como a estabilidade linearizada de Markov. O tempo de Markov t é equivalente a um parâmetro de resolução e a modularidade é recuperada em $t = 1$ para grafos não orientados (onde a notação *nor* traduz a ligação com o Laplaciano normalizado.).

$$r_{nor}^{lin}(1; H) = Q.$$

3.10 Módulo 9 - Sistemas Dinâmicos em e de Redes

3.10.1 Motivação

O estudo será sobretudo em situações “simples” que são analiticamente tratáveis, mas muitos dos processos dinâmicos que consideramos podem, obviamente, ser estudados em situações muito mais complicadas (inclusive em redes orientadas, redes pesadas, redes temporais e redes multicamadas), e muitos novos fenómenos interessantes ocorrem nessas situações que poderão ser alvo de trabalhos finais específicos por parte dos alunos.

A presença de dinâmica temporal de vértices e/ou arestas da rede evoca questões cruciais na modelação, nomeadamente, de fenómenos reais em que é necessário avaliar se devemos estudar um sistema dinâmico numa rede temporal, estudar apenas uma rede temporal, estudar um sistema dinâmico numa rede independente do tempo (ou num conjunto de tais redes) ou realizar vários estudos desse tipo.

A ênfase pode ser colocada na interação entre o local e a dinâmica global ou na topologia em evolução dinâmica. Mais interessante é uma combinação de ambos. Normalmente, tal combinação envolve um acoplamento entre duas escalas de tempo diferentes: uma rápida, na qual a dinâmica individual ocorre, e uma lenta, na qual a rede responde a essa dinâmica e evolui.

Nas redes neuronais, por exemplo, tem-se uma dinâmica de atividade rápida dos vértices, chamados neurónios neste contexto, e uma dinâmica de aprendizagem lenta que altera os pesos das conexões, chamadas sinapses, em resposta às correlações de atividade entre os neurónios.

3.10.2 Distinção entre modelos para a dinâmica nas redes e a dinâmica das redes

Os modelos para a dinâmica nas redes são a extensão mais natural dos modelos tradicionais de sistemas dinâmicos. Eles consideram como os estados das componentes, ou vértices, mudam ao longo do tempo através de suas interações com outros vértices que estão ligados a eles. As conexões são representadas por arestas de uma rede, onde a topologia da rede é fixa ao longo do tempo.

Nos modelos para a dinâmica das redes referimo-nos aos diferentes tipos dos chamados processos que ocorrem nas redes - a funcionalidade/eficiência de tais processos é fortemente afetada pela topologia, bem como pelo comportamento dinâmico da rede e a dinâmica das redes refere-se principalmente a vários fenômenos (por exemplo, auto-organização, agrupamento evolutivo) que ocorrem para provocar certas mudanças na topologia da rede.

3.10.3 Alguns sistemas dinâmicos estudados em redes.

3.10.4 Percolação

O estudo das mudanças qualitativas na conectividade em sistemas (especialmente os grandes) à medida que seus componentes são ocupados ou removidos, é um dos processos mais simples que apresenta fenômenos críticos ou transição de fase, ou seja, existe um parâmetro no sistema, em que uma pequena mudança produz uma grande mudança no comportamento do sistema. Imaginemos uma floresta entre duas estradas paralelas: nas margens de uma das estradas inicia-se um incêndio e com a “ajuda” do vento ocorre a propagação do fogo em direção à outra estrada. Quando as chamas atingirem a estrada no outro lado da mata, quanto da floresta já foi queimada?

Podemos ver na figura 3.2 que: em a) uma paisagem ocupada por 90% de células florestais (verde, $p=0.9$) distribuídas de forma aleatória e independente, dominadas por um grande aglomerado; em b) Para valores de p maiores que o limite de percolação ($p=0,6 > 0,59$) existe um caminho

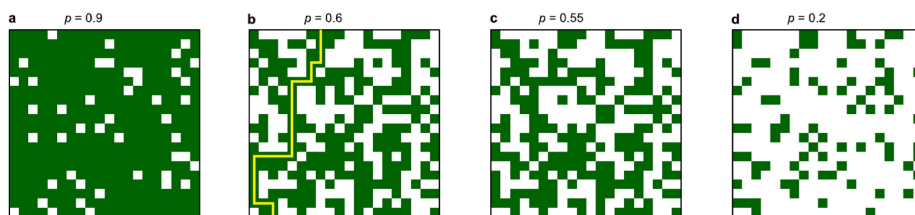


Figura 3.2: Exemplo de percolação na modelação da propagação do incêndio

contínuo a partir de duas bordas opostas (caminho amarelo dentro do cluster abrangente); em c) Abaixo do limiar de percolação (aqui $p=0,5 < 0,59$) surgem aglomerados maiores, mas nenhum aglomerado abrangente pode ser detectado e em d) uma paisagem ocupada por 20% de floresta apresenta pequenos aglomerados não ligados. O tamanho da paisagem é de 20x20 células, [Gra 22].

A percolação nos ajuda-nos a compreender a estrutura, a resiliência e os processos em redes complexas. As transições de percolação fornecem um arquétipo para transições contínuas, e tem havido muito trabalho em problemas de percolação. Muitos problemas de percolação estão profundamente relacionados com modelos tanto biológicos como sociais, como o modelo suscetível-infectado (SI) para contágios biológicos e o modelo de limite crítico de Watts para contágios sociais. Consideramos percolação nos vértices, nas arestas (removemos nodos, arestas até que uma componente gigante deixe de existir de forma uniforme ou não-uniforme) e percolação explosiva.

3.10.5 Contágios Biológicos

Uma das formas padrão de estudar contágios biológicos é através do que é tradicionalmente chamado de modelos compartimentais, nos quais os compartimentos descrevem um estado (por exemplo, “suscetível,” “infectado” ou “recuperado”) e existem parâmetros que representam taxas de transição para mudança de estado.

Os modelos compartimentais mais simples são aplicáveis a populações “bem misturadas”, nas quais cada indivíduo pode encontrar todos os outros indivíduos e em que cada tipo de mudança

de estado tem uma única probabilidade associada.

Modelo Suscetível-Infectado (SI):

O tipo mais simples de epidemia. biológica tem dois estados – suscetível e infectado – onde os nododulos saudáveis são considerados “suscetíveis” (e estão no compartimento “S”) porque não estão atualmente infectados, mas podem ser infectados, e os nodos “infectados” (no compartimento “I”) permanecem permanentemente nesse estado.

Modelo Suscetível-Infectado-Suscetível (SIS):

Consideramos um processo de transmissão de doença (um pouco mais complicado) no qual um nododulo pode tornar-se suscetível novamente após ser infectado.

Modelo Suscetível-Infectado-Recuperado (SIR):

No qual os nodos suscetíveis ainda podem fazer a transição para “ser” infectados, mas os nodos infectados recuperam para um estado R no qual não podem mais ser infectados.

3.10.6 Contágios Sociais

As ideias espalham-se pelas redes sociais de uma forma que parece ser algo análoga aos contágios biológicos, e a semelhança percebida entre as epidemias sociais e biológicas levou à adoção do termo “contágio” ao descrever a influência social e a propagação de ideias, inovações, e memes

3.10.7 Modelos de Votação

No modelo padrão (ou “direto”) cada nó está associado a uma variável binária que pode estar no estado $+1$ ou no estado -1 , (por exemplo, o primeiro pode representar o partido A, e o último pode representar o partido B). A cada passo de tempo discreto, um nodo (digamos, o nodo i) é selecionado uniforme e aleatoriamente, e o nodo i então adota o opinião s_j de um de seus vizinhos j

(que é selecionado uniforme e aleatoriamente entre todos os seus vizinhos $i's$). Se i e j já estivessem votando da mesma forma antes do intervalo de tempo, então nenhuma mudança ocorreria.

3.10.8 Acoplamento de osciladores

Um sistema dinâmico muito estudado, que associa cada oscilador a um nó de uma rede e que permite investigar como a conectividade não trivial afeta fenômenos coletivos como a sincronização. Modelamos a dinâmica nas redes com variáveis dinâmicas independentes $x_i, y_i, \dots$, em cada vértice i . As variáveis são acopladas apenas ao longo das arestas da rede. É estudado com mais detalhe este caso.

3.10.9 Sincronização completa de uma rede linear

Uma rede de N sistemas dinâmicos caóticos idênticos, x_i ($i = 1, \dots, N$), os nodos ou elementos da rede, pode ser descrita por

$$x_i(t+1) = f(x_i(t)) + \sum_{\substack{j=1 \\ j \neq i}}^N a_{ij} \cdot [f(x_j(t)) - f(x_i(t))]$$

onde $a_{ij} \in \mathbb{R}_0^+$ ($i, j = 1, \dots, N$) são as constantes-força-de-acoplamento.ou, duma forma equivalente, por

$$\vec{x}(t+1) = (I_N + A) \cdot \vec{f}(\vec{x}(t))$$

designando por A a matriz de acoplamento.

A definição de sincronização completa de um acoplamento pode estender-se de maneira simples a uma tal rede.

Definição: Dizemos que a rede admite uma sincronização completa se existir uma função $s(t)$ tal que $x_i(t) = s(t)$, $\forall i=1, \dots, N$, é solução da igualdade anterior.

Num acoplamento que admite sincronização completa, a solução completamente sincronizada só é exponencialmente estável para alguns valores da constante-força-de-acoplamento.

Definição.: Dizemos que a rede sincroniza completamente se existir uma função $s(t)$ tal que $x_i(t) = s(t)$, $\forall i=1, \dots, N$, é solução exponencialmente estável..

Apesar de ser muito importante, o estudo da sincronização não pode ser “restrito” à sincronização completa. A modelação realista de muitas grandes redes exige inevitavelmente que outros tipos de sincronização sejam considerados. como o caso da sincronização desfasada, onde as iterações dos sistemas dinâmicos acoplados assumem os mesmos valores, só que fazem-no em instantes t diferentes.

Estudo da estabilidade destas soluções considerando diversos tipos de acoplamento.

É exemplificado um outro tipo de acoplamento, o acoplamento por comando-total. Se à rede anterior acrescentarmos um nodo y , que designamos por ditador, ligando-o de forma unilateral a todos os elementos da rede e utilizando uma mesma constante de acoplamento c , a que chamamos constante-força-de-comando, obtemos uma nova rede descrita por:

$$\begin{cases} x_i(t+1) = f(x_i(t)) + \sum_{\substack{j=1 \\ j \neq i}}^N a_{ij} \cdot [f(x_j(t)) - f(x_i(t))] + \epsilon [f(y(t)) - f(x_i(t))] \\ y(t+1) = f(y(t)) \end{cases}$$

$$\begin{cases} x_i(t+1) = f(x_i(t)) + \sum_{\substack{j=1 \\ j \neq i}}^N a_{ij} \cdot [f(x_j(t)) - f(x_i(t))] + \epsilon [f(y(t)) - f(x_i(t))] \\ y(t+1) = f(y(t)) \end{cases}$$

ou, duma forma equivalente, por

$$\vec{x}_0(t+1) = (I_{N+1} + A_0) \cdot \vec{f}(\vec{x}_0(t))$$

$$\text{sendo } A_0 = \begin{bmatrix} A - \epsilon I_N & \epsilon \vec{1} \\ \vec{0}^T & 0 \end{bmatrix}.$$

Estudamos a possibilidade de este novo nodo impor o seu comportamento a todos os outros elementos da rede, i.e. a possibilidade desta nova rede sincronizar completamente.

3.10.10 Estudo da sincronizabilidade duma rede

Estudo dos efeitos da estrutura da rede, ou seja, a topologia do grafo, na sincronizabilidade da rede, através do estudo de alguns parâmetros do grafo, como a condutância e o coeficiente de agrupamento definindo o ponto de viragem do agrupamento para identificar a formação de clusters, como na figura 3.3.

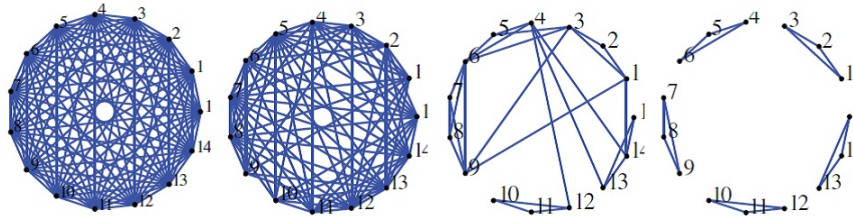


Figura 3.3: Processo de formação de cinco clusters.

Podemos observar o efeito na sincronização da rede, figura 3.4, considerando a evolução do coeficiente de agrupamento e da condutância no processo de formação do agrupamento C^5 , quando o número das arestas diminui.

3.10.11 Função zeta de um grafo

Ihara escreveu dois artigos nos anos noventa nos quais estabeleceu a estrutura para definir a função (que entretanto se chama Ihara) Ihara zeta de um grafo k -regular finito. A função zeta de um grafo

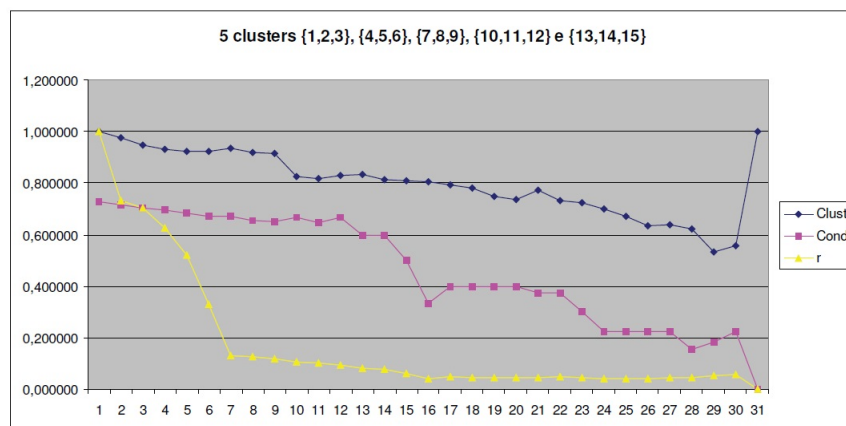


Figura 3.4: Evolução do coeficiente de agrupamento e da condutância. O eixo horizontal indica o número de etapas neste processo.

é o recíproco de um polinómio e pode ser calculado em tempo polinomial. Estudamos os coeficientes deste polinómio com o objectivo de os relacionar com a estrutura específica no grafo.

3.11 Bibliografia

Bibliografia

- [Ab 14] Abreu, N., Del-Vecchio, R., Trevisan, V., Vinagre, C. (2014). Teoria espectral de grafos - Uma introdução, III Colóquio de Matemática da Região Sul. Universidade Federal de Santa Catarina.
- [Alb-Bar 02] Albert, R., Barabási A., Statistical mechanics of complex networks, Rev Mod Phys 74, 47–97, 2002.
- [Adc 13] Adcock, A., Sullivan, B. and Mahoney, M., Tree-like structure in large social and information networks. In Data Mining (ICDM), IEEE 13th International Conference on Data Mining, 2013.
- [Ast 05] Aste, T., Di Matteo, T. and Hyde, S., Complex networks on hyperbolic surfaces, Physica A:Statistical Mechanics and its Applications 346, 20, 2005.
- [Bar-Alb, 99] Barabási, A. and Albert, R., Emergence of scaling in random networks, Science, 286(5439), 509–512, 1999.
- [Bar-Bia 23] Barabási, D., Bianconi, G., Bullmore, E. et al., Journal of Neuroscience 43, 5989, 2023.
- [Bat 13] Batty, M., The New Science of Cities, The MIT Press, Cambridge, MA., 2013.
- [Bea 83] Beardon, A., The Geometry of Discrete Groups, Springer-Verlag, 1983.

- [Bog 21] Marián Boguñá Dmitri Krioukov ^{1,2}, Ivan Bonamassa³, Manlio De Domenico and M. Ángeles Serrano, Network Geometry, Nat Rev Phys 3, 114–135, 2021.
- [Boll 02] Bollobas, B., Modern Graph Theory (Graduate Texts in Mathematics), Springer, 2002.
- [Bon 92] Bonacich, P. (1987). Power and centrality: a family of measures. American Journal of Sociology 92, 1170–1182.
- [Borg 07] Borgatti, S. P. (2007). Centrality and network flow. Social Networks, v. 27, 55–71.
- [Bran 05] Brandes U., Erlebach T., Network Analysis: Methodological Foundations, Vol. 3418 of Lecture Notes on Computer Science, Springer-Verlag, Berlin, 2005.
- [Caraça 76] Bento Jesus Caraça, A Cultura Integral do Indivíduo – problema central no nosso tempo, Separata da Gazeta de Matemática, 129-132, 1976.
- [Che 13] Chen, W., Fang, W., Hu, G. and Mahoney, M., On the hyperbolicity of small-world and 21 treelike random graphs, Internet Mathematics 9, 434, 2013.
- [Chu 97] Chung, F., Spectral graph theory, ser. Regional conference series in mathematics. Amer. Mathematical Society, no. 92, 1997.
- [Chu 06] Chung, F., Complex Graphs and Networks , CBMS Lecture Series, Number 107, AMS Publications, 2006,
- [Coo 05] Coolen, ACC., Kuehn, R. and Sollich, P., Theory of neural information processing systems, Oxford University Press, 2005.
- [Don 04] Donetti, L. and Munoz, M., Detecting network communities: a new systematic and efficient algorithm, Journal of Statistical Mechanics: Theory and Experiment P10012, 2004.

-
- [Dor 03] Dorogovtsev, S. and Mendes, J., Handbook of Graphs and Networks, chapter Accelerated Growth of Networks, Wiley-VCH, Berlin, 2003.
- [Emm 05] Emmer, M., C. Escher's Legacy : A Centennial Celebration, Springer-Verlag Berlin and Heidelberg, 2005.
- [Erd-Ren 59] Erdos, P. and Rényi, A. On Random Graphs I., Publicationes Mathematicae, 6, 290–297, 1959.
- [Erd-Ren 60] Erdos, P. and Rényi, A. On the evolution of random graphs., Publicationes Mathematicae, 5, 17–61, 1960.
- [For 10] Fortunato, S., Community detection in graphs. Physics reports, Elsevier, v. 486, n. 3, p. 75–174, 2010.
- [Fre 79] Freeman, L., Centrality in social networks: conceptual clarifications. Social networks 1, 215-239, 1979.
- [Gawe 13] Gawell, E., Non-euclidean Geometry in the Modeling of Contemporary Architectural Forms, The Journal of Polish Society for Geometry and Engineering Graphics, Volume 24, 2013.
- [God 01] Godsil, C., Royle, G., Algebraic Graph Theory. Springer, 2001.
- [Gra 02] Grácio, C., Sistemas Dinâmicos de Baixa Dimensão e Fluxo Geodésico no Bitoro, Tese de Doutorado, 2002.
- [Gra 10] Grácio, C. and Ramos, J.S., Geodesic length spectrum on compact Riemann surfaces, Journal of Geometry and Physics 60 11, 1643-1655, 2010.

- [Gra 14] C. Gracio, Coolen, ACC. and Annibale, A., Cycle statistics in complex networks and Ihara's Zeta function, Nonlinear maps and their applications, Springer Proceedings in Mathematics & Statistics 57, 81-94, 2014.
- [Gra 21] Martins, A., Grácio, C., Teixeira, C., Rodrigues, I., Zapata, JL. and Lúcia, F., Historia Augusta authorship: an approach based on Measurements of Complex Networks, Applied Network Science 6 1, 2021.
- [Gra 22] Perestrelo, S.; Grácio, C.; Ribeiro, N.; Lopes, L., A Multi-Scale Network with Percolation Model to Describe the Spreading of Forest Fires, Mathematics 2022.
- [Gro 87] Gromov, M., Hyperbolic groups Springer, New York, 1987.
- [Hadamard 1898] Hadamard, J., 1898, Les surfaces a courbures opposées et leurs lignes geodesiques, J. de Math. Pures et Appliquées, N°4, 27-73, 1898.
- [Hag 95] Hage, H. and Harary, F., Eccentricity and centrality in networks. Social networks, 17(1), 57-63, 1995.
- [Hua 23] Huang, W., Yu, Y., Xu, H., Su, Z. and Wu, Y., Hyperbolic Music Transformer for Structured Music Generation, in IEEE Access, vol. 11, 26893-26905, 2023.
- [Jos 10] Qian, M. Jost, J., Feng, J., Networks: From Biology to Theory, Springer, 2010.
- [Jos 15] Jost, J., Object oriented models vs. data analysis- is this the right alternative?, arXiv:1510.07172v1 [math.HO], 2015.
- [Jos 16] Jost, J., Mathematics as a Tool?, SFI Working Papers, 2016.
- [Kat 53] Katz, I. A new status index derived from sociometric analysis, Psychometrika, vol. 18, no. 1, 1953.

-
- [Krio, 10] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. and Boguna, M., Hyperbolic geometry of complex networks, *Physical Review E*, 82(3): 036106, 2010.
- [Lam 15] Lambiotte, R., Delvenne, J. and Barahona, M., Random walks, Markov processes and the multiscale modular organization of complex networks. *Trans. Netw. Sci. Eng.* 1(2), 76–90, 2015.
- [Les 09] Leskovec, J., Lang, K., Dasgupta, A. and Mahoney, M., Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, *Internet Mathematics* 6, 29, 2009.
- [Men 03] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks*, Oxford University Press, Oxford, 2003.
- [Moh 91] Mohar B, *The Laplacian Spectrum of Graphs, Graph Theory, Combinatorics and Applications*, Vol.2, 1991.
- [New 03] Newman, M, *Structure and function of complex networks*. *SIAM Review*, 45(2):167–256, 2003.
- [New 00] Newman, M, *Networks: An Introduction*, Oxford University Press, Oxford (2010).
- [New 06] Newman, M., Modularity and community structure in networks, *Proc. Natl. Acad. Sci. USA*, vol. 103, 8577–8582, 2006.
- [Pag-Bri 98] , Sergey Brin, Lawrence Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Comput. Networks* 30(1-7): 107-117 (1998).
- [Poincaré 1882] Poincaré, H., Théorie des groupes fuchsien, In: *Acta Math.* 1.1, First article in a series published in the same volumen, 1882.

- [Poincaré 1902] Poincaré, H., *La Science et l'hypothèse*. Paris, Flammarion, 1902.
- [Por 09] Porter, M., Onnela, J. and Mucha, P., *Communities in networks*, Notices of the AMS, vol. 56, no. 9, 1082–1097, 2009.
- [Por 16] Porter, M. and James Gleeson, J., *Dynamical Systems on Networks - a Tutorial*, Springer, 2016.
- [Scu 03] S. Bornholdt, S. and Schuster, G., editors. *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, 2003.
- [SR 82] Sousa Ramos, J., *A árvore das dinâmicas e soluções simbólicas de equações diferenciais*, C.F.M.C., E22, pg 1-17.
- [SR 89] Sousa Ramos, J., *Hiperbolicidade e Bifurcação de Sistemas Simbólicos*, Tese de Doutorado. 1989.
- [Ste 10] Steen, M., *Graph Theory and Complex Networks: An Introduction*, 144.1, 2010.
- [Wat-Str 98] Watts D. and Strogatz S., *Collective Dynamics of Small World Networks.*, Nature, 393, 440-442, 1998.
- [Str 06] Strogatz, S., *Exploring complex networks*, Nature, 410:268–276, 2006.
- [Wan 21] Wan, Z., Mahajan, Y., Kang, B., Moore, T. and Cho, J., *A Survey on Centrality Metrics and Their Network Resilience Analysis*, in IEEE Access, vol. 9, 104773-104819, 2021.

Apêndice A

Módulo 1 - Apresentação e Motivação

A.0.1 Motivação



Figura A.1: No festival Maha Kumbh Mela, na Índia, este ano, são esperados mais de 400 milhões de pessoas.

HUMANIDADE

Como nos organizamos???

As cidades surgiram como centro das civilizações e acompanharam a história humana desde a antiguidade até aos dias atuais

Em 2000, metade da população mundial vivia em cidades e as Nações Unidas projectam que atinja 60% em 2030 e que, para o ano 2050, 70% da população mundial seja urbana.



Figura A.2: As cidades surgiram como centro das civilizações e acompanharam a história humana desde a antiguidade até aos dias atuais.

Peter Grindrod (Professor Mathematical Institute, University of Oxford, ex-presidente do Instituto de Matemática e suas Aplicações), Desmond J. Higham (Professor da School of Mathematics, University of Edinburgh, Membro da Royal Society de Edimburgo), Robert S. MacKay (Professor de Matemática, University of Warwick, Membro da Royal Society).

“...No preâmbulo de seu recente livro *The New Science of Cities*, Michael Batty, do Bartlett Centre for Advanced Spatial Analysis da University College London, discute três princípios centrais

onde explica a sua perspectiva de “redes e fluxos” da ciência da cidade. O livro de Batty faz uso de conceitos como modelação baseada em agentes, teoria dos grafos, cadeias de Markov, problemas de decisão markovianos, otimização e auto-similaridade/fractais e, portanto, é um excelente ponto de partida para matemáticos que desejam entrar na área. . . ”

O desafio de grandes conjuntos de dados!



Figura A.3: Brennan Whitfield , Aug 07, 2023.

Somos confrontados com uma explosão nos últimos anos de dados biomédicos, como dados experimentais de processos bioquímicos e estruturas em níveis celulares, sub-celulares e até moleculares, as topologias de complexos sistemas compostos de sinalização e processamento de informações, como o cérebro ou o sistema imunológico, dados genômicos e epigenéticos (níveis de expressão genética, sequências de DNA), dados epidemiológicos e um grande número de registros de pacientes com informações clínicas.

Tende-se a recolher dados por um motivo. Esse motivo geralmente é o desejo de compreender o comportamento dinâmico do sistema complexo que gerou os dados, para prever com razoável precisão a sua evolução futura e a sua resposta a perturbações ou intervenções, ou para entender como se formou. Podemos querer melhorar e otimizar a eficiência de um sistema, para projetar controles regulatórios eficazes ou (no caso da medicina) para compreender, curar e prevenir doenças.

Para sistemas pequenos e simples a tradução da observação em compreensão qualitativa e quantitativa do design e a função geralmente não é difícil.

No entanto, se recolhermos dados sobre sistemas complexos com milhões de variáveis que intera-

gem de forma não linear, apenas tendo uma lista das suas partes, das suas conexões, as observações de seus padrões coletivos de comportamento já não são suficientes para entender como esses sistemas funcionam.

A partir de 2022, de facto, aproximadamente 97 por cento das empresas estão a investir no poder crescente do big data.

A matemática na vanguarda do conhecimento!

(Jürgen Jost, diretor do Max Planck Institute for Mathematics in the Sciences in Leipzig desde 1996)

“...Portanto, como a recolha de dados hoje em dia está tipicamente à frente da compreensão teórica, a matemática deveria enfrentar radicalmente a falta de teoria e olhar para o que existe, os dados, e ver o que se pode fazer com eles. E aquilo a que pretendo apelar não são métodos ad-hoc para cada conjunto de dados concretos, mas sim uma análise abstrata dos desafios estruturais mais profundos.

É claro que a matemática moderna está desenvolvida e sofisticada o suficiente para fornecer ferramentas apropriadas e úteis para basicamente qualquer conjunto de dados, mas esta, por si só, é uma perspetiva científica demasiado estreita.

Para mim, como matemático, a matemática é mais do que análise de dados. Precisamos de um repensar conceitual.

Uma perspetiva tradicional consideraria a matemática apenas como uma ferramenta quando os conjuntos de dados (possivelmente muito grandes) fossem analisados. Uma das teses deste ensaio é que esse próprio processo, a análise de dados, pode e já se tornou objeto de pesquisa matemática. Assim, a matemática não só serve como uma ferramenta poderosa, mas ao refletir esse papel, ganha uma nova perspetiva e se eleva a um nível superior de abstração. Então, o domínio de tal investigação matemática não é mais um campo específico, como a física, mas a própria técnica matemática. A matemática deixa então de ser, se é que alguma vez foi, uma mera ferramenta formal para a ciência, mas torna-se a ciência das ferramentas formais...”

O século da complexidade

A percepção de que estamos inseridos e cercados por diversas redes e que estas possuem um papel central em diversos aspectos de nossas vidas, vem crescendo vertiginosamente durante a última década, tanto no meio acadêmico quanto no público em geral. Estamos cada vez mais cientes de que o comportamento de muitas das coisas que nos cercam não pode ser estudado e caracterizado isoladamente, pois muitas destas coisas estão conectadas e a interação entre as partes influencia fundamentalmente o comportamento individual e coletivamente o comportamento global. O que torna estas redes complexas é que são geralmente tão grandes que é impossível compreender ou prever o seu comportamento global observando o comportamento de nodos ou ligações individuais

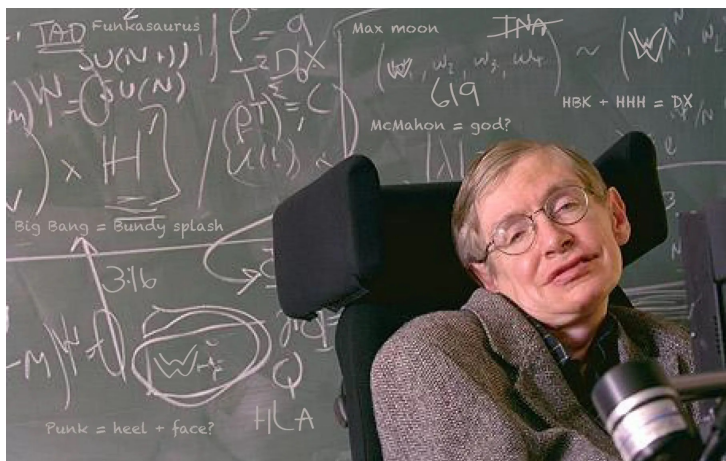


Figura A.4: Stephen Hawking afirmou em 2000 que "... O século XXI será o século da Complexidade" e que "...A teoria definitiva não colocará limites na complexidade dos sistemas que podemos produzir, e é nessa complexidade que penso que os desenvolvimentos mais importantes do próximo milênio ocorrerão..."@

Figura A.4: Stephen Hawking afirmou em 2000 que "... O século XXI será o século da Complexidade" e que "...A teoria defi

Assim sendo, surge naturalmente a necessidade de estudar como as “entidades” se conectam e a importância desta conectividade para o problema em questão.

Algum contexto histórico:

1736- Euler introduz a ideia de grafo (na sua busca por um caminho ligando sete pontes em Königsberg).

1950- Freeman – medidas de centralidade.

1957- Redes aleatórias por Salomon e Rapoport.

1959-1960 -Redes aleatórias por Erdos e Renyi.

1965- Price propõe modelo matemático ligação preferencial (Herbet Simon).

1967- Stanley Milgram e o Mundo Pequeno com 6 graus de separação.

1998 – Modelo Watson-Strogatz para mundo pequeno e Page Rank (Google).

1999- Modelo Albert-Barabasi introduz as redes livre de escala.

Logo a teoria dos grafos tem uma história que remonta a mais de 250 anos e desde então, a teoria dos grafos, o estudo das redes na sua forma mais básica como interconexões entre os objetos, evoluiu de suas raízes recreativas para um ambiente rico e assunto distinto.

De particular importância é o seu papel vital na nossa compreensão da matemática que governa o universo discreto.

A mesma abstração é usada para representar a conectividade da Internet e a conectividade do cérebro

O que é surpreendente é que muitas redes do mundo real são parecidas: a estrutura da Internet assemelha-se à organização do nosso cérebro, mas também à organização das comunidades sociais online.

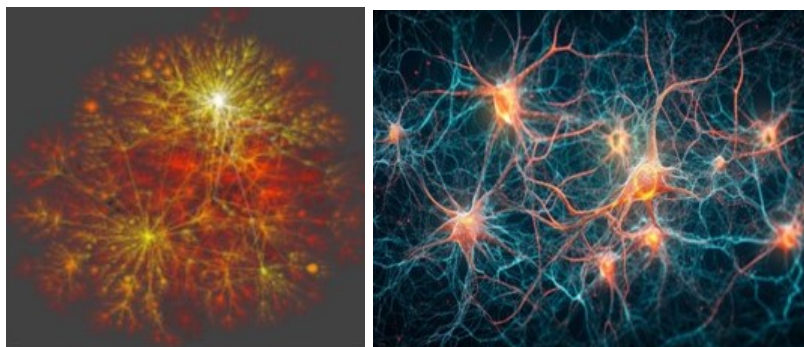


Figura A.5: Do lado esquerdo uma representação gráfica da internet: Os nodos representam as páginas da web e os links representam instruções de salto em HTML. Do lado direito estão representadas as sinapses no nosso cérebro.

A Internet é um exemplo do que hoje é comumente chamado de rede complexa, que podemos definir informalmente como uma grande coleção de nodos interligados.

O cérebro pode ser visto como uma rede onde os neurónios são nodos e suas sinapses são as arestas.

Semelhança entre redes imunológicas, neuronais e circuitos

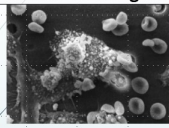
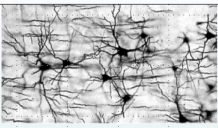
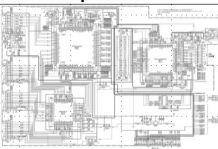
redes imunológicas	redes neuronais	computadores
		
10 ⁸ concentrações de clones B/T hora links paralelos adaptativos baixa conectividade equações?	10 ¹¹ sequências de picos de neurónios por msecs links paralelos adaptativos alta conectividade equações conhecidas desde 1940/1950	10 ¹⁰ portas lógicas 0/1 estados nsecs links fixos sequenciais baixa conectividade equações conhecidas desde 1940/1950

Figura A.6: Comparação de redes.

A origem dessas semelhanças ainda é um mistério, assim como muitas vezes é muito difícil compreender como certas redes foram realmente estruturadas

O que torna estas redes complexas é que são geralmente tão grandes que é impossível compreender ou prever o seu comportamento global observando o comportamento de nodos ou ligações individuais.

Mesmo grafo, diferentes redes

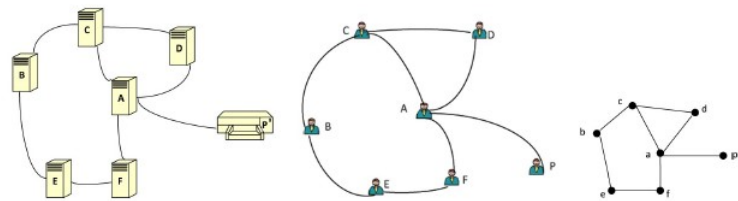


Figura A.7: Do lado direito da figura uma rede de routers, ao centro uma rede de amigos e no lado esquerdo o grafo que representa as duas redes.

Nem todo grafo é uma rede complexa, mas toda a rede complexa pode ser representada por um grafo

As redes estão em toda a parte - a ubiquidade da teoria de redes

Um vértice pode ser qualquer coisa: uma pessoa, uma organização, um computador, uma flor ou uma célula biológica, etc.

Interligados significa que dois nodos podem estar ligados, por exemplo, porque duas pessoas se conhecem, duas organizações trocam bens, dois computadores têm um cabo conectando os dois, ou porque dois neurónios estão ligados por meio de sinapses para a passagem de sinais.

Acontece que redes complexas estão por toda parte. Ou, para ser mais preciso, acontece que se modelarmos situações do mundo real em termos de redes, muitas vezes descobrimos coisas novas.

Antes de aprofundarmos que as redes complexas realmente implicam, vamos primeiro considerar algumas áreas gerais onde as redes desempenham um papel vital, redes tecnológicas (de comunicação, de computadores, Internet), redes biológicas (moleculares, neuronais, metabólicas), redes de texto, redes de números, redes sociais, redes colaborativas...

Rede de estradas e ruas

Rede das estradas romanas

Roma era famosa pelo seu sistema de estradas. Os romanos construíram mais de 85 mil quilómetros de estradas para ligar todas as partes do seu império.



Figura A.8: Rede de estradas romanas.

Não é por acaso que existe o ditado: “Todos os caminhos levam a Roma”.

Rede das ruas no centro histórico de Évora

Podemos considerar uma fotografia do Centro Histórico de Évora, o mapa desenhado das suas ruas e o grafo que representa essa rede de ruas.



Figura A.9: Podemos considerar uma fotografia do Centro Histórico de Évora, o mapa desenhado das suas ruas e o grafo que representa essa rede de ruas.

Redes de transporte



Figura A.10: Rede do metropolitano de Lisboa e rede de tráfego aéreo.

Rede Sísmica Europeia Virtual de Banda Larga

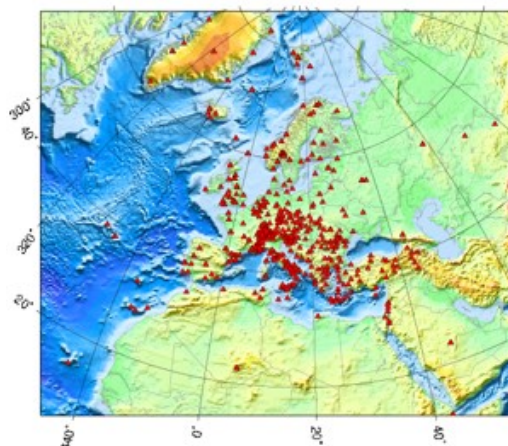


Figura A.11: Rede Sísmica.

Redes Sociais

Rede facebook

O Facebook é a rede social mais popular do mundo, com cerca de 2900 milhões de nodos por mês.

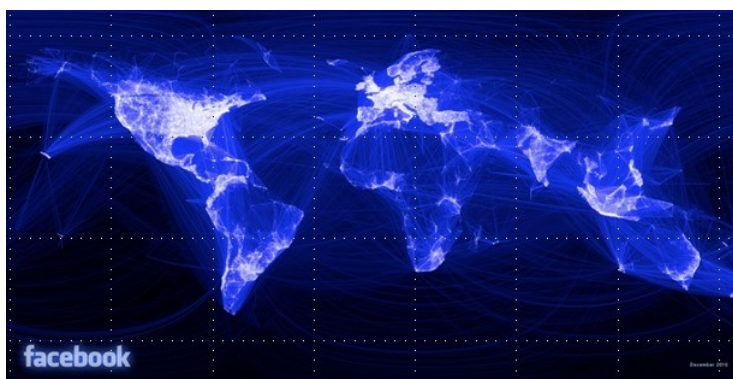


Figura A.12: Rede Facebook.

Rede Twitter

A rede Twitter tem cerca de 436 milhões de participantes por mês.

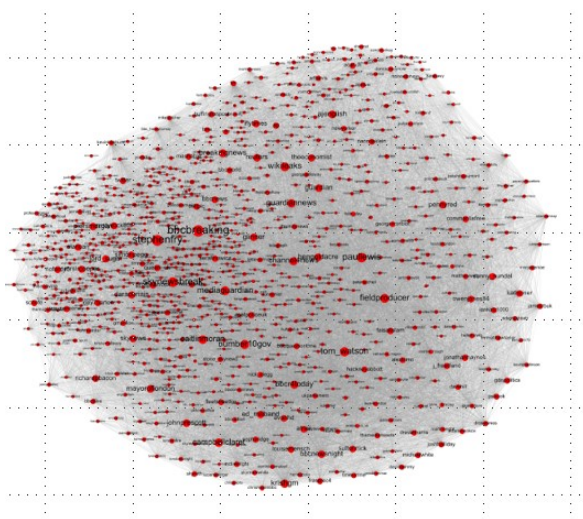


Figura A.13: Rede Twitter.

Redes profissionais

Rede de colaboração científica

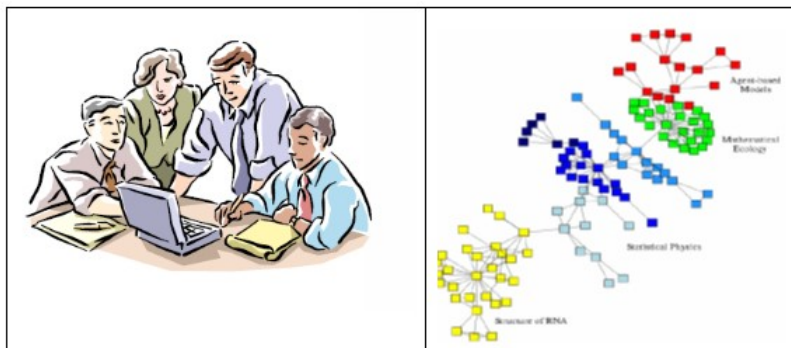


Figura A.14: Uma rede de colaboração científica.

Rede de alunos que frequentam os mesmos cursos

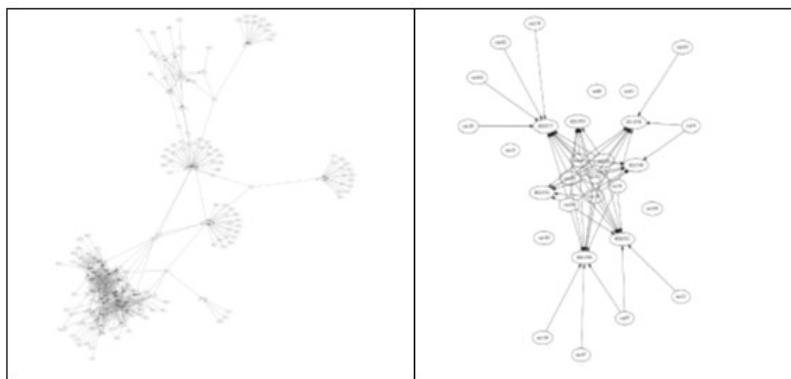


Figura A.15: uma rede formada pelos alunos e pelas disciplinas nos cursos em e-learning da Universidade de Évora.

Redes de texto

Rede de co-ocorrências num livro

“História Augusta”: coleção de biografias dos imperadores romanos do período 117 a 284, escrita em latim. Cada livro é dividido em partes mais curtas do texto, com o mesmo número de palavras e é construída uma rede de co-ocorrências

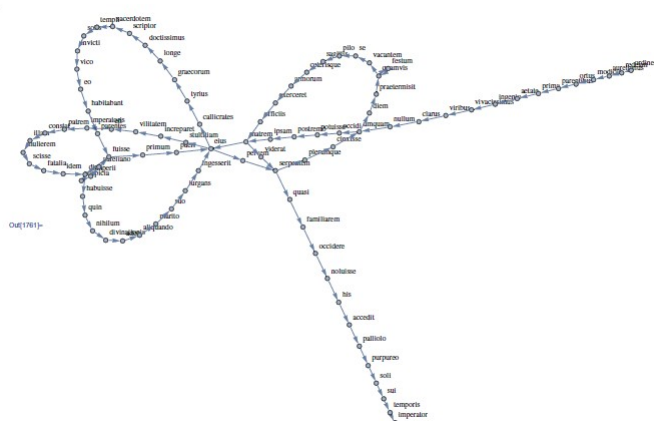


Figura A.16: Rede de co-ocorrências.

Rede de Personagens de Harry Potter

Neste caso constrói-se uma rede com as personagens que se cruzam nestes livros.

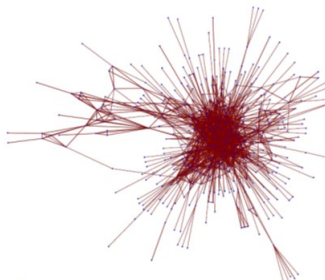


Figura A.17: Rede de Personagens de Harry Potter.

Rede de Números

Também podemos considerar uma rede de números, neste caso consideramos a sequência de Collatz.

$$C(x) = \begin{cases} 3x + 1 & \text{se } x \equiv 1(mod 2), \\ \frac{x}{2} & \text{se } x \equiv 0(mod 2). \end{cases}$$



Figura A.18: Rede de números da sequência de Collatz.

Redes biológicas

Rede neuronal de *Caenorhabditis elegans*

Caenorhabditis elegans, mais conhecidos como *C. elegans*, são vermes microscópicos que normalmente crescem até 1 mm de comprimento. *C. elegans* foi isolado pela primeira vez em 1900 e, desde o final da década de 1960, tem sido usado para “modelar” doenças humanas. Isso ocorre porque *C. elegans* e os humanos partilham algumas características fisiológicas comuns e têm uma sobreposição significativa nos seus códigos genéticos. A simplicidade do sistema nervoso de um *C. elegans* (é composto apenas por algumas centenas de neurônios enquanto a rede neuronal humana tem na ordem de 100 bilhões de neurônios e 100 trilhões de sinapses), tem sido uma das razões para ser muito utilizado no desenvolvimento de modelos para várias doenças, nomeadamente neurodegenerativas.

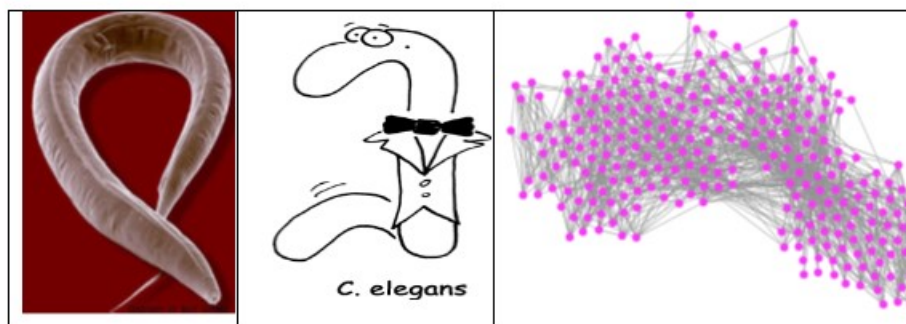


Figura A.19: A rede neuronal do nematóide *C. elegans*.

Perguntas fundamentais

Os primórdios do campo da “ciência das redes”, que pode ser caracterizada como a ciência da conectividade, começou por se focar na estrutura da rede.

Este estudo é valioso, mas também é preciso considerar a dinâmica, e é bom usar um chapéu dinâmico mesmo para investigações cujo foco principal é a estrutura. Quando estudamos um processo dinâmico, preocupamo-nos com o seu comportamento como função do tempo, espaço e seus parâmetros.

Existem numerosos estudos que examinam quantas pessoas estão infectadas por um contágio biológico e se ele persiste de uma estação para outra. Todos esses estudos têm algo em comum: a dinâmica está ocorrendo num conjunto de entidades discretas (os nodos numa rede) que estão conectadas umas às outras através de arestas de alguma maneira não trivial.

Leva à pergunta natural de como a conectividade não trivial subjacente afeta os processos dinâmicos.

Esta é uma das questões mais importantes em ciência de redes. É importante realçar que não só a estrutura da rede pode afetar a dinâmica dos processos numa rede, mas também que os processos dinâmicos podem afetar a própria dinâmica da rede.

Por exemplo, quando uma criança fica com gripe, ela pode não ir à escola por alguns dias, e essa mudança temporária na atividade humana afeta quais os contatos sociais que ocorrem, o que pode, por sua vez, afetar a dinâmica da propagação da doença.

Perguntas fundamentais para as quais esta disciplina deverá fornecer ferramentas:

- Que leis governam/explicam o funcionamento desses sistemas formados por componentes que interagem entre si?
- Como modelar o comportamento desses sistemas?
- Entender a estrutura de conexões (topologia), estudar a dinâmica de evolução, criar algoritmos para resolver problemas
- Identificar agrupamentos (comunidades), vértices críticos.

A consideração explícita do tempo abre um conjunto inteiramente novo e rico de questões de investigação, tais como:

- Como as redes de inovação se formam e evoluem? Que mecanismos são responsáveis pelas suas características?
- Como as redes afetam a estabilidade dos sistemas dos quais fazem parte?
- Quais são as consequências da formação ou remoção de entradas ou saídas de organizações na rede para a estabilidade estrutural do sistema?
- Quais os processos que podem ser usados para modelar a dinâmica?

Estas duas últimas perguntas colocam questões diferentes:

- **Dinâmicas em redes:** mudanças dinâmicas de estado que ocorrem numa topologia de rede estática
- **Dinâmica de redes:** crescimento dinâmico e transformação de topologias de rede
(Estes são os modelos que consideram mudanças da própria topologia de rede ao longo do tempo)

Algumas considerações

Como vimos as redes são interessantes tanto matematicamente quanto computacionalmente, e estão difundidas na física, na biologia, na sociologia, na ciência da informação e numa infinidade de outros campos.

O estudo de redes é uma das “estrelas em ascensão” do conhecimento científico, e as redes tornaram-se num dos assuntos mais importantes para os matemáticos aplicados estudarem. A maioria dos tópicos a serem considerados são áreas ativas de pesquisa moderna.

Espero que, nesta disciplina, além de desenvolverem um sólido conhecimento e apreciação de algumas das ferramentas, conceitos e cálculos utilizados no estudo de redes, desenvolvam, também, a capacidade de ler e compreender artigos de investigação atuais na área.

A.0.2 Programa

Conteúdos programáticos da ficha curricular da disciplina:

- A teoria de redes dinâmicas é uma combinação da teoria de grafos e da dinâmica não-linear
- Teoria qualitativa dos grafos
- Compreensão da teoria básica de sistemas dinâmicos caóticos
- Relação entre a dinâmica e a estrutura da rede
- Relação entre a dinâmica local e global
- Sincronização de redes
- Análise espectral de redes
- Propriedades específicas de diferentes tipos de redes

A.0.3 Funcionamento da disciplina

Aulas teórico-práticas com apresentação teórica dos tópicos abordados e discussão de exemplos e casos de estudo.

Resolução de pequenos problemas com aplicação das metodologias selecionadas e uso de software existente. Implementação de algoritmos selecionados. Desenvolvimento de um trabalho final.

São colocados na plataforma moodle material de apoio para cada módulo incluindo listas de exercícios.

Em cada módulo haverá um fórum de discussão semanal, onde o aluno deverá colocar questões sobre a matéria e sobre os exercícios propostos.

Programação semanal

- 1. Apresentação e motivação (1 semana)
- 2. Conceitos básicos (2 semanas)
- 3. Teoria espectral (1 semana)
- 4. Medidas de centralidade (2 semanas)
- 5. Redes Aleatórias e modelos de configuração (2 semanas)
- 6. Redes “small world” e redes “Watt-Strogatz”(1 semana)
- 7. Redes Sem escala (1 semana)
- 8. Partição de redes e detecção de comunidades (1 semana)

- 9. Dinâmicas em Redes e de Redes (3 semanas)
- 10. Apresentação dos trabalhos Finais (1 semana)

A.0.4 Avaliação

1ª Parte

Resolução, entrega e discussão de um certo número de exercícios das listas de exercícios fornecidas. A essa resolução será dada uma nota, a Nota TE.

2ª Parte

No decorrer dos últimos módulos, cada aluno escolhe um tema para um miniprojecto individual, o trabalho final (TF), que desenvolverá e enviará por escrito.

O trabalho deverá depois ser discutido através da plataforma Zoom. A esse trabalho e à sua apresentação será dada uma nota, a Nota TF.

A nota final será calculada pela fórmula: $NF = NE(30\%) + TF(70\%)$

Apêndice B

Notas Introdutórias

A Ciência pode ser encarada sob dois aspectos diferentes. Ou se olha para ela tal como vem exposta nos livros de ensino, como coisa criada, e o aspecto é o de um todo harmonioso, onde os capítulos se encadeiam em ordem, sem contradições. Ou se procura acompanhá-la no seu desenvolvimento progressivo, assistir à maneira como foi sendo elaborada, e o aspecto é totalmente diferente - descobrem-se hesitações, dúvidas, contradições, que só um longo trabalho de reflexão e apuramento consegue eliminar, para que logo surjam outras hesitações, outras dúvidas, outras contradições.

Descobre-se ainda qualquer coisa mais importante e mais interessante: - no primeiro aspecto, a Ciência parece *bastar-se a si própria*, a formação dos conceitos e das teorias parece obedecer só a necessidades interiores; no segundo, pelo contrário, vê-se toda a influência que o ambiente da vida social exerce sobre a criação da Ciência.

A Ciência, encarada assim, aparece-nos como um *organismo vivo*, impregnado de *condição humana*, com as suas forças e as suas fraquezas e subordinado às grandes necessidades do homem na sua luta pelo *entendimento* e pela *libertação*; aparece-nos, enfim, como um grande capítulo da vida humana social.

Será esta a atitude que tomaremos aqui. A Matemática é geralmente considerada como uma ciência à parte, desligada da realidade, vivendo na penumbra do gabinete, um gabinete fechado, onde não entram os ruídos do mundo exterior, nem o sol nem os clamores dos homens. Isto, só em parte, é verdadeiro.

Sem dúvida, a Matemática possui *problemas próprios*, que não têm ligação imediata com os outros problemas da vida social. Mas não há dúvida também de que os seus fundamentos mergulham *tanto como os de outro qualquer ramo da Ciência*, na vida real; uns e outros entroncam na mesma *madre*.

Bento de Jesus Caraça.

Conceitos Fundamentais da Matemática, 1941.

B.1 Introdução

A teoria dos grafos e o seu ensino é um assunto antigo e consolidado. Também é possível encontrar cursos sobre diversos aspetos de redes em departamentos como estatística, ciência da computação, sociologia, entre outros. Muitos deles já existem há algum tempo, mas a noção de estudar a matemática das redes - envolvendo assuntos como a teoria dos grafos, mas distinta dela em aspetos cruciais - é relativamente nova, e é necessário que os currículos de matemática nos cursos de graduação e pós-graduação incluam estes assuntos. A importância de lecionar tais disciplinas vai muito além do estabelecimento do tema “redes” como tendo uma identidade distinta de disciplinas como a teoria dos grafos. O estudo de dados discretos passou por uma revolução e é necessário que as pessoas com formação em matemática tenham conhecimento destes conceitos. Tendo em conta a diversidade de conhecimentos dos alunos o presente texto, sempre em atualização, introduz e/ou relembra conceitos iniciais de teoria de grafos e de redes com o objectivo de servir de apoio ao estudo de Redes Dinâmicas, disciplina semestral optativa do plano de Doutoramento em Matemática.

O desafio de grandes conjuntos de dados!

Nas últimas décadas, a nossa capacidade de coleccionar e armazenar vastas quantidade de dados aumentou dramaticamente. Isso inclui dados socioeconómicos, ligações sociais entre indivíduos ou redes de colaboração profissional; dados de tráfego em sistemas informáticos e de telecomunicações, redes de satélites, internet, redes eléctricas, ligações ferroviárias, rodoviárias ou aéreas e redes de distribuição.

Recolhemos e processamos grandes quantidades de dados geológicos e meteorológicos, dados sobre o nível do mar, poluição do ar e da água, registos vulcânicos e sísmicos e tamanhos de camadas de gelo polar e glacial. Finalmente, vimos uma explosão nos últimos anos de dados biomédicos, como dados experimentais sobre bioquímica processos e estruturas nos níveis celular, subcelular e mesmo molecular, as topologias de sistemas compostos complexos de sinalização e processamento de informações, como o cérebro ou o sistema imunológico, dados epigenéticos, dados epidemiológicos e grande número de registos de pacientes com informações clínicas.

Mas tem que existir um motivo para guardar este mundo de dados. Esse motivo geralmente é o desejo de compreender o comportamento dinâmico do sistema complexo que gerou os dados, para prever com uma razoável precisão a sua evolução futura e sua resposta a perturbações ou intervenções, ou para entender como se formou.

Um primeiro e útil passo na modelação de dados estruturais em sistemas complexos de muitas variáveis é visualizar esses sistemas como redes.

O objectivo de nossa disciplina é estudar redes dinâmicas.

A teoria de redes dinâmicas está preocupada com sistemas de unidades acopladas de acordo com uma estrutura gráfica subjacente. Por isso, investiga a interação entre dinâmica e estrutura,

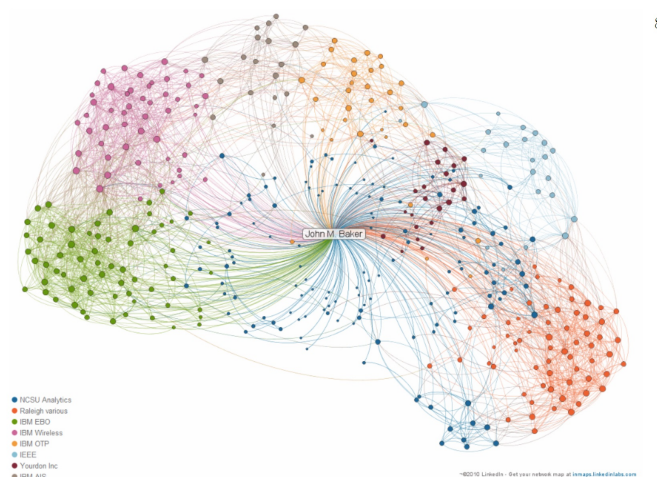


Figura B.1: Rede de colaboradores da IBM.

entre os processos temporais acontecendo nas unidades individuais e a estrutura espacial estática que as liga.

Quando estudamos um processo dinâmico, preocupamo-nos com o seu comportamento como função do tempo, espaço e seus parâmetros. Existem numerosos estudos que examinam quantas pessoas estão infectadas por um contágio biológico e se ele persiste de uma estação para outra. Todos esses estudos têm algo em comum: a dinâmica está ocorrendo em um conjunto de entidades discretas (os nodos numa rede) que estão conectadas umas às outras através de arestas de alguma maneira não trivial. Isso leva à pergunta natural de como a conectividade não trivial subjacente afeta os processos dinâmicos. Este é uma das questões mais importantes em ciência de redes. É importante realçar que não só a estrutura da rede pode afetar a dinâmica dos processos numa rede, mas também que os processos dinâmicos podem afetar a própria dinâmica da rede. Por exemplo, quando uma criança fica com gripe, ela pode não ir à escola por alguns dias, e essa mudança temporária na atividade humana afeta quais os contatos sociais que ocorrem, o que pode, por sua vez, afetar a dinâmica da propagação da doença.

No século XVIII, a cidade de Königsberg foi dividida por um rio, chamado Pregel. Continha

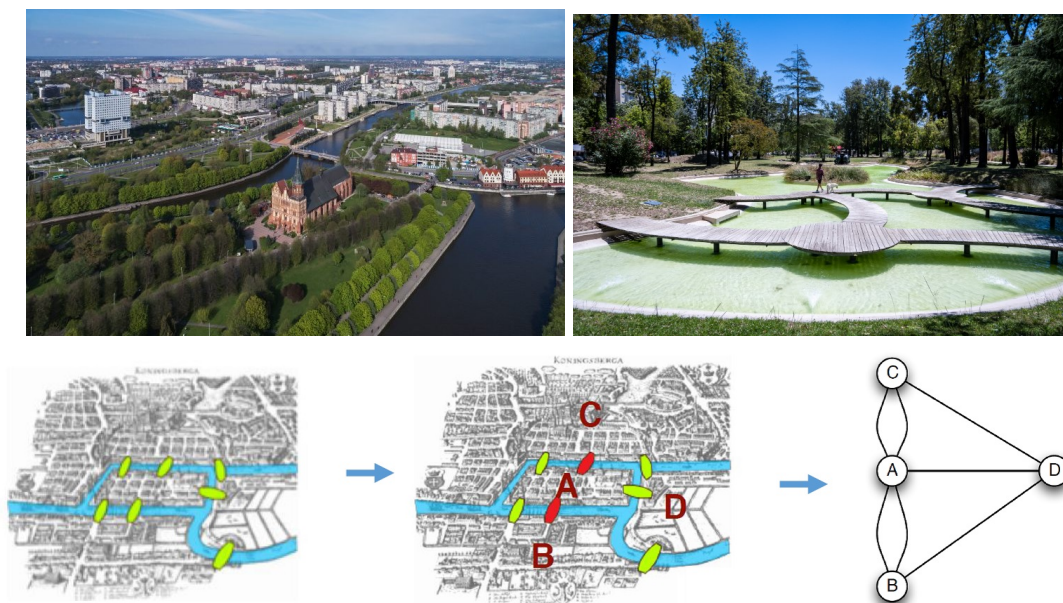


Figura B.2: Na parte superior temos no lado esquerdo as pontes de Kaliningrado (antiga Königsberg) e no lado direito as sete pontes que encerram em si o problema matemático no jardim do Campo Grande. Na parte inferior temos o estudo do problema das pontes de Euler.

duas ilhas e sete pontes ligando as várias partes do território. Um quebra-cabeça famoso na época era encontrar um passeio pela cidade que atravessasse cada ponte exatamente uma vez. Muitas pessoas afirmaram ter encontrado tal caminhada, mas quando solicitadas a reproduzi-la, ninguém conseguiu. Em 1736, o matemático Leonhard Euler explicou o porquê: ele mostrou que tal caminho não existia. A solução de Euler é surpreendentemente simples — basta que se veja o problema da maneira correta. O truque é cortar todas as informações desnecessárias.

O resultado de Euler marcou o início da **teoria dos grafos**, o estudo de redes feitas de pontos ligados por linhas. Desde então, a teoria dos grafos evoluiu das suas raízes recreativas para um campo riquíssimo e com um objecto distinto.

Como podemos observar na Figura B.3, o mesmo grafo, com 7 vértices e 8 arestas, representa duas redes diferentes, a rede de computadores e a rede de amigos.

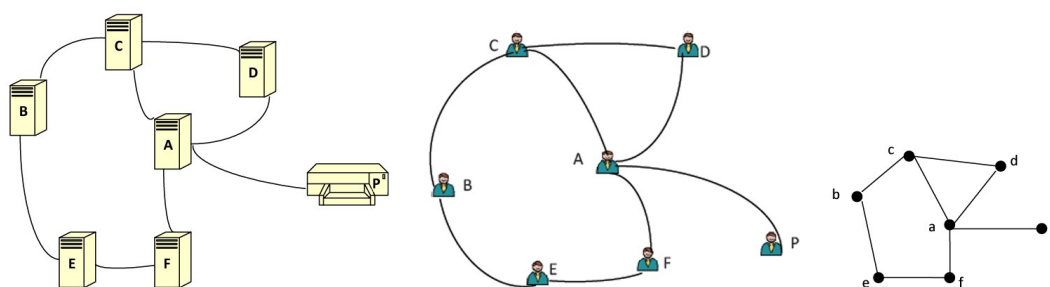


Figura B.3: No lado esquerdo uma rede de computadores, ao centro uma rede de amigos e do lado direito um grafo que representa as duas redes.

A matemática na vanguarda do conhecimento!

Hoje, muitas áreas consideram a ciência das redes como sua. Matemáticos reivindicam corretamente a propriedade e a prioridade por meio da teoria dos grafos; a exploração de redes sociais por sociólogos remonta a décadas; a física emprestou o conceito de universalidade e infundiu muitas ferramentas analíticas que agora são inevitáveis no estudo das redes; a biologia investiu centenas de milhões de euros no mapeamento de redes subcelulares; a ciência da computação ofereceu perspectiva algorítmica, permitindo explorar redes muito grandes; a engenharia investiu esforços consideráveis na exploração de infraestruturas redes. É notável como tantas peças díspares conseguiram "encaixar-se", dando origem a uma nova disciplina.

Vejamos o que afirma, numa entrevista, o conhecido matemático Jürgen Jost cuja área científica primária é a geometria.

Portanto, como a recolha de dados hoje em dia está tipicamente à frente da compreensão teórica, a matemática deveria enfrentar radicalmente a falta de teoria e olhar para o que existe, os dados, e ver o que se pode fazer com eles. E aquilo que pretendo apelar não são métodos ad-hoc para cada conjunto de dados concretos, mas sim uma análise abstrata dos desafios estruturais mais profundos. É claro que a matemática moderna está desenvolvida e sofisticada o suficiente para fornecer ferramentas apropriadas e úteis para basicamente qualquer conjunto de dados, mas

esta, por si só, é uma perspectiva científica demasiado estreita. Para mim, como matemático, a matemática é mais do que análise de dados. Precisamos de um repensar conceptual. Uma perspectiva tradicional consideraria a matemática apenas como uma ferramenta quando conjuntos de dados (possivelmente muito grandes) fossem analisados. Uma das teses deste ensaio é que esse próprio processo, a análise de dados, pode e já se tornou objeto de pesquisa matemática. Assim, a matemática não só serve como uma ferramenta poderosa, mas ao refletir esse papel, ganha uma nova perspectiva e se eleva a um nível superior de abstração. Então, o domínio de tal investigação matemática não é mais um campo específico, como a física, mas a própria técnica matemática. A matemática deixa então de ser, se é que alguma vez foi, uma mera ferramenta formal para a ciência, mas torna-se a ciência das ferramentas formais

B.2 Definições

As redes que foram introduzidas até agora são matematicamente conhecidos como grafos. Na sua forma mais simples, um grafo é uma coleção de vértices que podem ser ligados uns aos outros por meio de arestas. Em particular, cada aresta do grafo une exatamente dois vértices.

B.2.1 Tipos de grafos

Definição 1.1: Um **grafo** G consiste numa coleção V de vértices e numa coleção de arestas E , para as quais escrevemos $G = (V, E)$. Diz-se que cada aresta $e \in E$ une dois vértices. Se e une $u, v \in V$, escrevemos $e = \langle u, v \rangle$. Os vértices u e v neste caso são ditos adjacentes. A aresta e é dita incidente com vértices u e v , respectivamente

É importante perceber que não se faz distinção entre $\langle v, u \rangle$ e $\langle u, v \rangle$, ambos representam o facto de que os vértices u e v são adjacentes. Podemos perguntar se é possível que uma aresta una os mesmos vértices, ou seja, uma aresta pode formar um laço? Não há nada na definição que impeça isso e, de facto, tais arestas são autorizadas. Da mesma forma, podemos perguntar se é possível

ligar dois vértices u e v por múltiplas arestas? Na verdade, isso também é possível e um grafo que não tem laços, nem arestas múltiplas é chamado **grafo simples**.

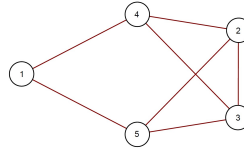


Figura B.4: Grafo simples.

Da mesma forma, não há nada que proíba um grafo de não ter vértices o que implica que, nesse caso, também não haverá arestas. Chamar-se-á um grafo vazio. Outro caso especial é formado por um grafo simples com n vértices, em que cada vértice é adjacente a todos os outros vértices. Este grafo é conhecido por **grafo completo**. Um grafo completo com N vértices é comumente denotado como K_N . Excluindo laços, temos $\frac{N(N-1)}{2}$ arestas. Diz-se que o grafo está totalmente conectado.

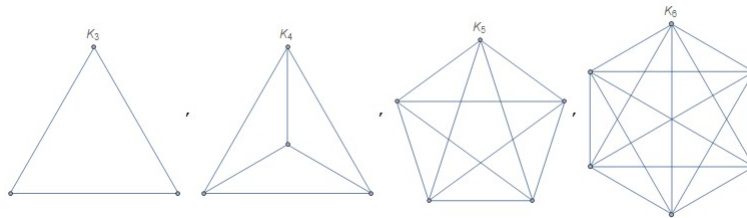


Figura B.5: Grafos completos K_3 , K_4 , K_5 e K_6 .

Para avaliarmos "até que ponto um grafo está quase completo ou não" podemos considerar a sua densidade, o número de arestas de um grafo dividido pelo número máximo possível de arestas.

Definição 1.2: Considere um grafo simples G com N vértices e m arestas. A **densidade** $\rho(G)$ de G é definida por: $\rho(G) = \frac{m}{\binom{N}{2}}$.

Um grafo diz-se **denso** se a sua densidade é alta, isto é próxima de 1. Um grafo diz-se **esparso** se a sua densidade é baixa, isto é próxima de 0.

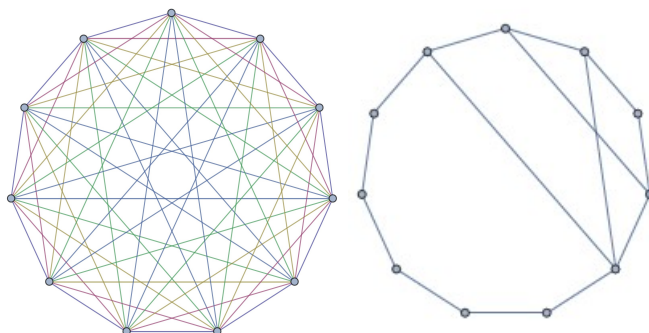


Figura B.6: No lado esquerdo um grafo denso e no lado direito um grafo esparso com o mesmo número de vértices.

Definição 1.2: Para qualquer grafo G e vértice $v \in V(G)$, o conjunto dos **vizinhos** $N(v)$ de v é o conjunto dos vértices (diferentes de v) adjacente a v :

$$N(v) \stackrel{\text{def}}{=} \{w \in V(G) : v \neq w, \exists e \in E(G) : e = \langle u, v \rangle\}.$$

Definição 1.3 Seja G um grafo simples, o seu **complementar** designa-se por $\overline{G}$ e contém todas as arestas que não estão em G .

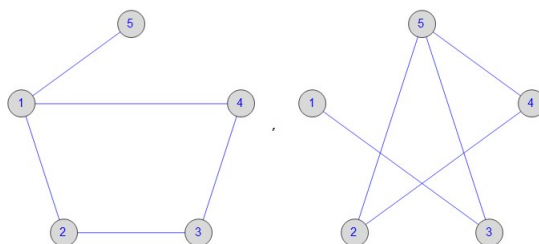


Figura B.7: Grafo G e o seu complementar.

Definição 1.3 Um grafo **planar** é um grafo onde existe pelo menos uma representação gráfica no plano do grafo sem o cruzamento de arestas.

Definição 1.3: Denomina-se **ordem** de G a cardinalidade do seu conjunto de vértices, $V(G)$ e tamanho de G a cardinalidade do seu conjunto de arestas, $E(G)$.

Outro conceito importante de grafos é o de subgrafo.

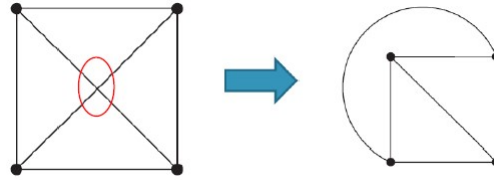
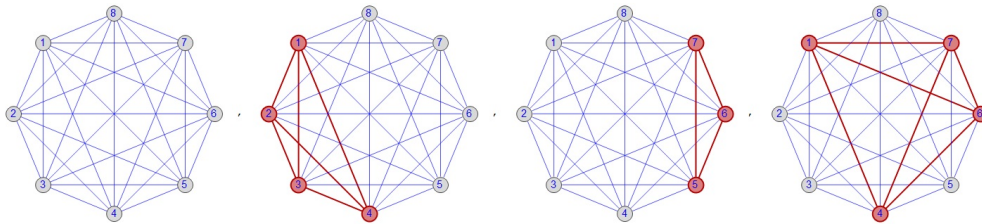


Figura B.8: Grafo planar.

Definição 1.4: Um grafo H é um **subgrafo** de G se $V(H) \subseteq V(G)$ e $E(H) \subseteq E(G)$ tal que para todo $e \in E(H)$ com $e = \langle u, v \rangle$, temos que $u, v \in V(H)$. Quando H é um subgrafo de G , escrevemos $H \subseteq G$.

Figura B.9: O grafo G e alguns subgrafos.

Definição 1.5: Um **clique** num grafo G é um subconjunto máximo $V' \subseteq V$ de vértices no grafo tal que cada membro do subconjunto tenha uma aresta que se liga a todos os outros vértices. Aqui 'máximo' significa que é impossível adicionar qualquer outra aresta a V' tal que a nova aresta se conecte a todas as arestas anteriores em V' .

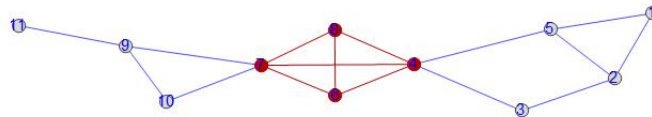


Figura B.10: Está desenhado um clique no grafo.

Definição 1.6: Considere um grafo simples $G = (V, E)$. O **grafo linha** de G , denotado como $L(G)$ é construído a partir de G representando cada aresta $e = \langle u, v \rangle$ de E por um vértice v_e em

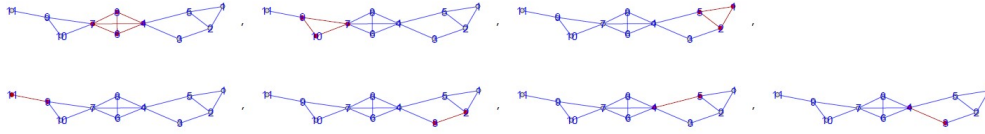


Figura B.11: Estão representados todos os cliques do grafo.

$L(G)$, e juntando dois vértices v_e e v_{e^*} se e somente se as arestas e e e^* são incidentes com o mesmo vértice em G .

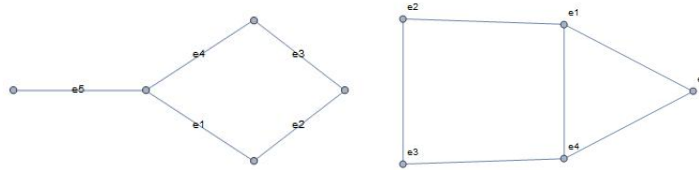


Figura B.12: No lado esquerdo o grafo G e no lado direito o seu grafo linha $L(G)$.

No caso mais simples, um grafo é simétrico e sem pesos, e há uma ligação entre os elementos i e j quando eles interagem. Numa evolução dinâmica, a influência não precisa ser simétrica, ou seja, o estado de i pode ser uma entrada para o cálculo do próximo estado de j , mas não inversamente.

Definição 1.7: Chama-se grafo **orientado** (ou **digrafo**) quando existem ligações não simétricas, ou seja, $\exists(i, j) \in E$ tal que $(j, i) \notin E$. Neste caso a aresta (j, i) é desenhado com uma seta do nodo i para o nodo j .

Na tabela da Figura B.14 elencamos algumas redes reais distinguindo se se trata dum grafo ou dum digrafo.

Como vimos numa evolução dinâmica, a interação pode não ser simétrica. Mas, também, pode ocorrer com uma determinado força w_{ji} que se codifica atribuindo o peso w_{ji} a aresta de i a j . Obtemos assim um grafo pesado.

Por exemplo, ao considerar uma rede ferroviária como um grafo, as estações ferroviárias são

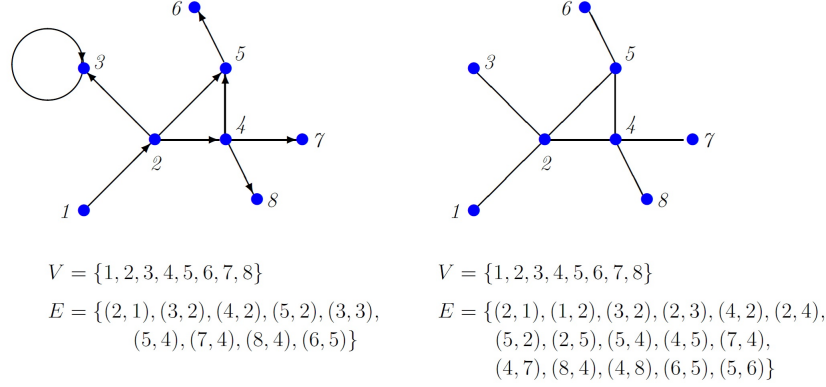


Figura B.13: À direita está desenhado um grafo e à esquerda está desenhado um grafo orientado.

naturalmente representadas por vértices, enquanto duas estações adjacentes são conectadas por meio de uma aresta. Em seguida, atribuímos um peso a uma aresta que representa a distância entre essas duas estações.

Definição 1.9: Chama-se **grafo pesado** G quando a cada aresta $e \in E(G)$ esteja associado um número real $w(e)$. Ou seja, uma aresta pesada entre os nodos v_i e v_j é $e = (v_i, v_j, w)$, onde w é o peso da aresta.

Definição 1.10.: Um grafo bipartido G é aquele em que os vértices podem ser divididos em dois subconjuntos disjuntos não vazios, ou seja, $V = V1 \cup V2$ com $|V1|, |V2| > 0$ e $V1 \cap V2 = \emptyset$, tal que todas as arestas $(i, j) \in E$ têm ou $i \in V1$ e $j \in V2$ ou $j \in V1$ e $i \in V2$.

Um exemplo duma rede bipartida é a rede formada por investigadores e revistas científicas onde publicam.

Podemos ainda considerar um grafo no qual um par de vértices pode ter mais de uma aresta uni-los. Basta pensar no caso das pontes da cidade de Königsberg estudado por Euler. Neste caso dizemos que se trata dum multigrafo. Podemos observar na Figura B.15 um grafo bipartido e um multigrafo.

Redes	Vértices	Arestas	<u>Grafo</u> Digrafo	N	L	$\langle k \rangle$
Internet	Routers	Conexões	Grafo	192244	609066	6.34
WWW	Páginas web	Links	Digrafo	325729	1497134	4.6
Rede Energética	Transformadores	Cabos	Grafo	4941	6594	2.67
Chamadas, T. portateis	Subscritores	Chamadas	Digrafo	36595	91826	2.51
Rede, de, E-mail	Endereços	E-mails	Digrafo	57194	103731	1.81
Colaboradores, científicos	Cientistas	Co-autoria	Grafo	23133	93439	8.08
Rede, de, actores	Actores	co-participação	Grafo	702388	29397908	83.71
Rede, de, citações	Artigos	Citações	Digrafo	449673	4689479	10.43
Metabolismo, E. Coli	Metabólitos	Reações químicas	Digrafo	1039	5802	5.58
Interações, proteínas	Proteínas	Interação, vinculativa	Grafo	2018	2930	2.90

Figura B.14: Tabela com várias redes complexas indicando o tipo de grafo, os seus vértices, as suas arestas e o grau médio.

Definição 1.11: Diz-se que $G = (V; E)$ é um **multigrafo** se existirem duas arestas e_1 e e_2 em E e se existirem dois vértices v_1 e v_2 em V tal que $e_1 = \langle v_1, v_2 \rangle$ e $e_2 = \langle v_1, v_2 \rangle$.

Definição: Considere dois grafos $G = (V, E)$ e $G^* = (V^*, E^*)$. G e G^* são **isomórficos** se existir uma aplicação injectiva $f : V \rightarrow V^*$ tal que para cada aresta $e \in E$ com $e = \langle u, v \rangle$, existe uma única aresta $e^* \in E^*$ com $e^* = \langle f(u), f(v) \rangle$.

B.2.2 A matriz de adjacência e de incidência de um grafo

Os grafos podem ser desenhados de diferentes formas, mas ao considerar a sua definição formal, os grafos são meramente descritos em termos de vértices e arestas. A questão da apresentação dos grafos é particularmente importante quando precisamos representar e estudar grafos muito grandes.

Num grafo de N -vértices existem $N \times N$ potenciais arestas, a presença de cada um dos quais pode ser codificada por um número binário, que organizamos como as entradas de uma matriz.

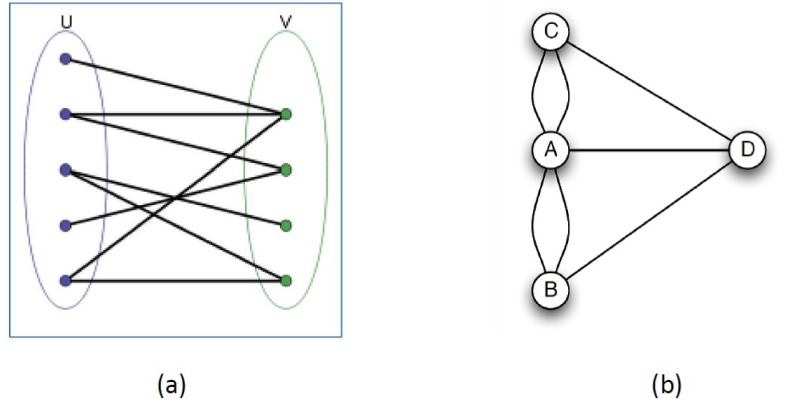


Figura B.15: (a) Grafo bipartido ; (b) Multigrafo.

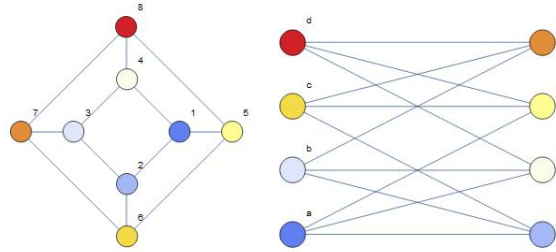


Figura B.16: Dois grafos isomorfos onde a mesma cor identifica os vértices correspondentes .

Definição 1.8 : A matriz de adjacência $A \in \{0,1\}^{N \times N}$ de um grafo com N -vértices $G = (V; E)$ é definida por:

$$\forall (i, j) \in \{1, \dots, N\}^2 : \begin{cases} A_{ij} = 1 & \text{se } (i, j) \in E, \text{ i.e se, existir um link } j \rightarrow i \\ A_{ij} = 0 & \text{se } (i, j) \notin E, \text{ i.e se não existir link } j \rightarrow i \end{cases}$$

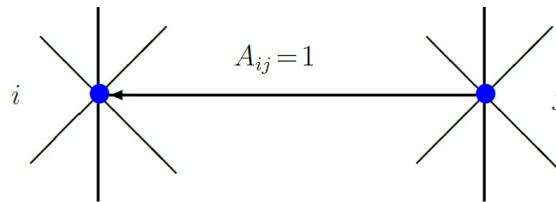


Figura B.17: Entrada A_{ij} da matriz.

Existe uma correspondência unívoca entre as entradas binárias de A e a especificação de quais

links estão presentes num grafo, logo cada grafo corresponde a uma matriz de adjacência única e vice-versa.

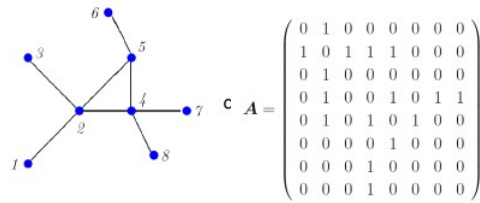


Figura B.18: Matriz de adjacência de um grafo.

Pdemos observar que no primeiro caso, Figura B.18, um grafo simples, a **matriz de adjacência** é **simétrica** e tem zeros na diagonal (não tem laços) enquanto no segundo caso, um grafo orientado com laços a matriz já não é simétrica e tem um elemento não nulo na diagonal.

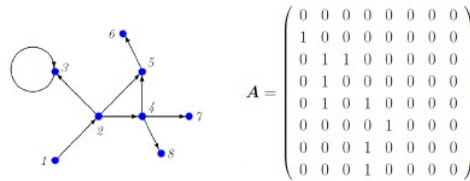


Figura B.19: Matriz de adjacência de um digrafo.

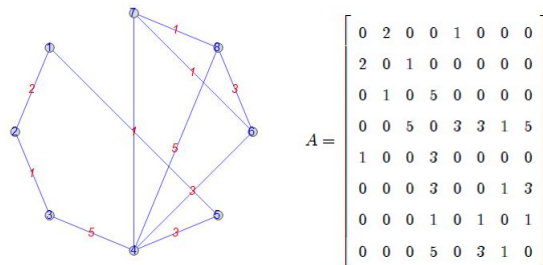


Figura B.20: Um grafo pesado e a sua matriz de adjacência.

No caso em que o grafo é pesado (orientado ou não) em vez de 1 na entrada da matriz será w_{ji} na interseção da linha i -ésima e a coluna j se existir um link de i para j .

Consideremos um grafo (ou digrafo) G com a matriz de adjacência A . Com esta matriz podemos calcular o número de arestas L .

Para um grafo simples $L = \sum_{i < j} A_{ij}$, e para um grafo orientado $L = \sum_{i < j} A_{ij} + \sum_i A_{ii}$.

O cálculo da densidade também se torna mais simples.

$$\rho(G) = \frac{\sum_{i < j} A_{ij}}{\frac{1}{2}N(N-1)}, \text{ para um grafo simples.}$$

$$\rho(G) = \frac{\sum_{i < j} A_{ij} + \sum_i A_{ii}}{\frac{1}{2}N(N-1) + N} = \frac{\sum_{i < j} A_{ij} + \sum_i A_{ii}}{\frac{1}{2}N(N+1)}, \text{ para um grafo.}$$

$$\rho(G) = \frac{\sum_{ij} A_{ij}}{N^2}, \text{ para um grafo orientado.}$$

Como alternativa, também podemos representar um grafo através da sua **matriz de incidência**. Uma matriz de incidência M do grafo G tem n linhas e m colunas tais que $M[i, j]$ conta o número de vezes que a aresta e_j é incidente com o vértice v_i . Notemos que $M[i, j]$ é 0, 1 ou 2. Consideremos o exemplo representado na Figura B.21.

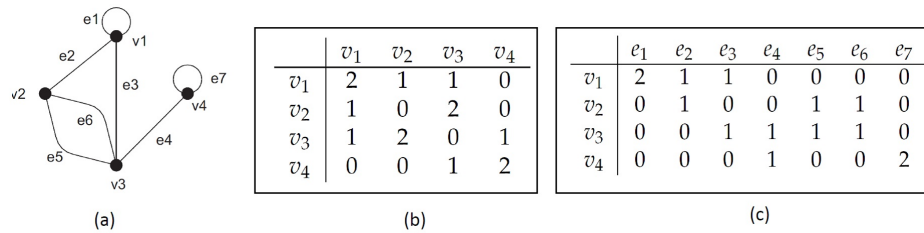


Figura B.21: (a) Grafo G , (b) matriz de adjacência de G e (c) matriz de incidência de G .

B.2.3 Grau de um vértice

Uma propriedade chave de cada vértice é o seu grau, representando, por exemplo, o número de contactos telefónicos que um indivíduo tem na rede de chamadas telefónicas (ou seja, o número de indivíduos com quem a pessoa conversou), ou o número de citações que um artigo científico tem numa rede de citações..

Definição 1.4: O número de arestas associadas a um vértice (nodo) v_i é chamado o **grau** de v_i (ou centralidade de grau).

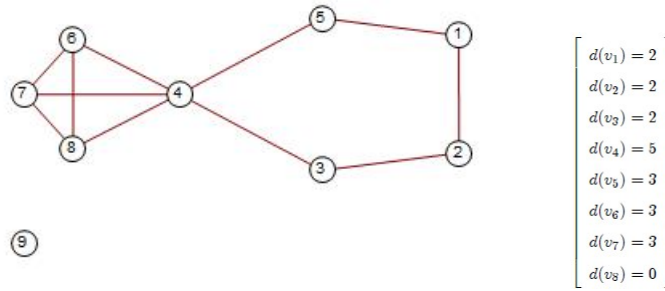


Figura B.22: Grafo G e os graus dos vértices.

Na Figura B.22 o vértice 9 não tem nenhuma aresta associada logo o seu grau é zero enquanto que o vértice 4, que tem 5 arestas associadas, o seu grau é 5.

Se todos os vértices tiverem o mesmo grau, o grafo é chamado **grafo regular**. Num grafo k -regular cada vértice tem grau k . Chamam-se grafos cúbicos o caso especial dos grafos 3- regulares.

Podemos calcular facilmente o grau de um vértice v_i , através da matriz de adjacência, a soma dos valores na linha i é igual ao grau do vértice v_i .

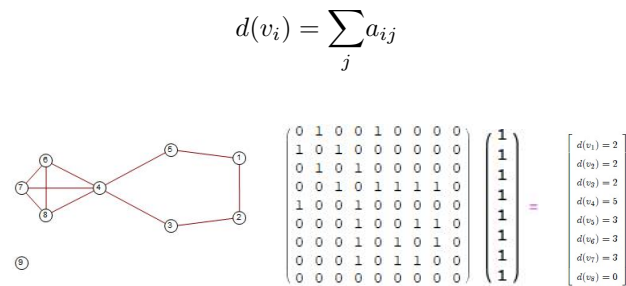


Figura B.23: Na última coluna temos os graus dos vértices.

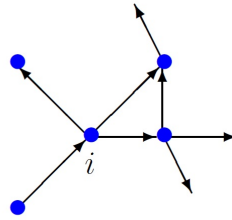
Teorema 1.1: Dado um grafo G , a soma dos graus dos vértices é o dobro do número de arestas,

ou seja, $\sum_{v \in V(G)} d(v) = 2|E(G)|$.

No exemplo anterior, da Figura B.23, temos que $\sum_{i=1}^6 d(v_i) = 14$ e $2|E(G)| = 2 \times 7 = 14$.

Para redes representadas por grafos orientados, temos que considerar o grau de saída, que representa a quantidade de arestas que partem de um vértice, e o grau de entrada, que corresponde ao número de arestas que tem o vértice em questão como destino. O grau total é definido como a soma dos dois anteriores. Se considerarmos o vértice i temos que:

$$k_i^{out} = \sum_j a_{ji}; \quad k_i^{in} = \sum_j a_{ij}; \quad k_i = k_i^{out} + k_i^{in} = \sum_j a_{ji} + \sum_j a_{ij}.$$



$$\begin{aligned} k_i^{in}(\mathbf{A}) &= 1 \\ k_i^{out}(\mathbf{A}) &= 3 \end{aligned}$$

Figura B.24: Grau $k_i = 4$ do vértice i .

Se listarmos os graus dos vértices $\{1, 2, \dots, N\}$ de um grafo G , com matriz de adjacência A , obtemos uma sequência de graus.

$$(k_1(A), k_2(A), \dots, k_N(A)).$$

Definição: O **grau médio** de um grafo de N -vértices com matriz de adjacência A é dado por:

$$\bar{k}(A) = \langle k \rangle = \frac{\sum_{i=1}^N k_i}{N}.$$

Claro que se considerarmos um grafo orientado temos que calcular o grau médio de saída e o grau médio de entrada. Na Figura B.14 estão calculados os graus médios de algumas redes reais.

Definição: a) O **grau máximo** é igual ao valor de grau mais alto alcançado por qualquer vértice.

$$d_{\max} = \max_{u \in V} d(u).$$

b) O **grau mínimo** é igual ao valor de grau mais baixo alcançado por qualquer vértice.

$$d_{\min} = \min_{u \in U} d(u).$$

Podemos normalizar o grau máximo e o grau mínimo dividindo o seu valor pelo grau médio.

O número total de arestas também pode ser calculado através da soma dos graus dos vértices.

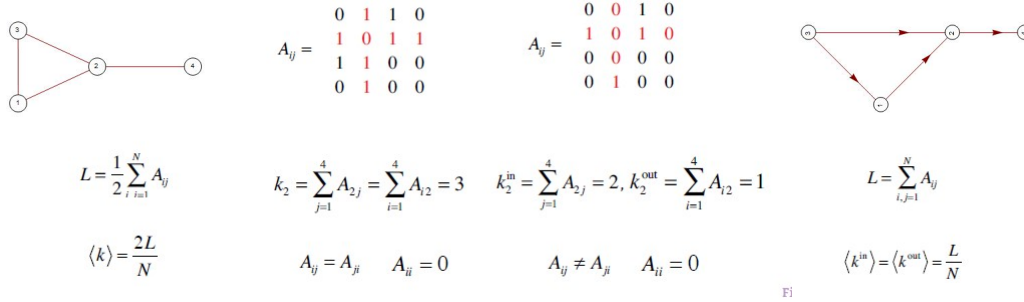


Figura B.25: Exemplo do cálculo dos graus, do número de arestas e do grau médio, num grafo (lado esquerdo) e num grafo orientado (lado direito).

B.2.4 Distribuição dos graus

A seqüência dos graus pode ser representada através dum histograma. Nesse caso, para um grafo simples com n vértices, escrevemos o valor $h(k)$ como sendo o número de vértices de grau k .

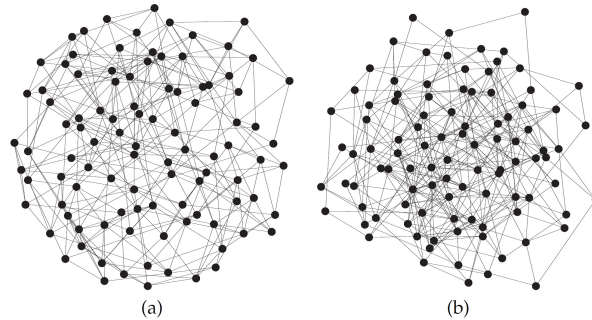


Figura B.26: Dois grafos (a) e (b) diferentes mas com o mesmo número de vértices.

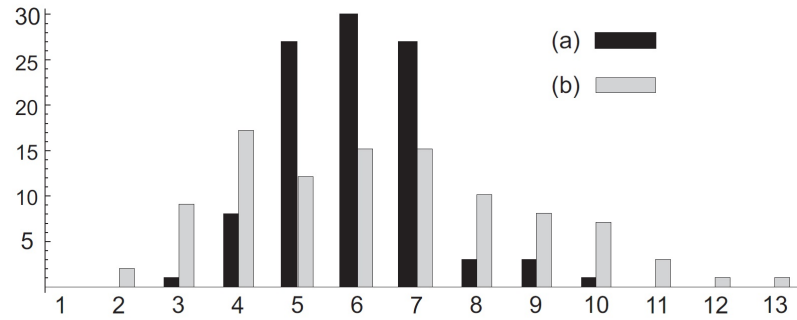


Figura B.27: Histograma com as distribuições dos graus dos vértices dos grafos (a) e (b).

Estes dois grafos parecem semelhantes mas, no entanto, ao considerar as suas respectivas distribuições de graus, vemos que estamos, de facto, em presença de dois grafos muito diferentes. Para completar esta análise simples, notamos que ambos os grafos possuem 100 vértices, mas que o grafo (a) tem 300 arestas, e o grafo (b) com 317 arestas.

A distribuição de graus, p_k , fornece a probabilidade de um vértice no grafo ter grau k . Como p_k é uma probabilidade, deve ser normalizada: $\sum_{k=1}^{\infty} p_k = 1$.

A distribuição de graus assume um papel central na teoria das redes e o cálculo da maioria das propriedades de rede exige que conheçamos p_k .

Para uma rede com N nodos a **distribuição dos graus** é dado por:

$$p_k = \frac{N_k}{N}$$

onde N_k é o número de nodos de grau k .

Usando a distribuição de graus de G , podemos obter outra expressão para o grau médio de G .

Definição: A **distribuição de graus** de um grafo simples com N vértices, com matriz de adjacência A é definida por:

$$p(k|A) = \frac{1}{N} \sum_i \delta_{k, k_i(A)}, \forall k \in \mathbb{N}.$$

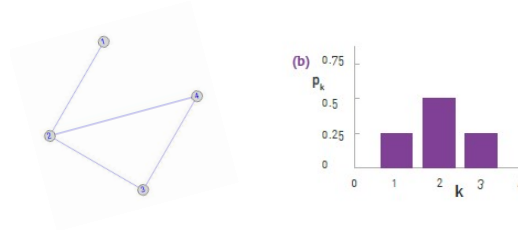


Figura B.28: $p_1 = 1/4$ (um dos quatro nodos tem grau $k_1 = 1$), $p_2 = 2/4 = 1/2$ (dois nodos têm $k_3 = k_4 = 2$), e $p_3 = 1/4$ (um dos quatro nodos tem grau $k_2 = 3$) e $p_k = 0$ para qualquer $k > 3$ (não existem nodos com grau $k > 3$).

Representa para cada k a fração de vértices i no grafo que possuem grau $k_i(A) = k$.

É notável que este conceito seja o mais simples e mais estudado que podemos encontrar numa rede e é usado como base para a caracterização da topologia da rede.

Uma distribuição de graus também pode ser dada a partir de uma fórmula (por exemplo, distribuição de Poisson ou outra).

Em redes reais, os graus dos nodos podem variar imenso. Para uma melhor leitura a distribuição de graus é representada, frequentemente, usando uma escala logarítmica.

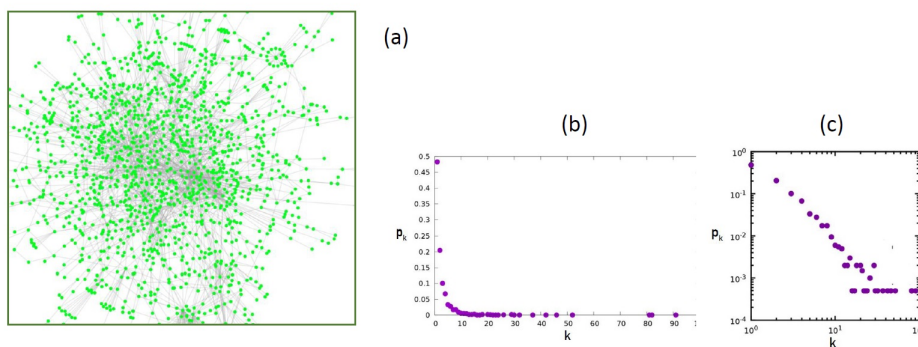


Figura B.29: (a) rede de interação de proteínas; (b) distribuição dos graus; (c) Distribuição dos graus utilizando uma escala logarítmica.

Na figura Fig. B.29 podemos observar que os graus variam entre $k = 0$ (nodos isolados) e $k = 92$, (este vértice com o maior grau chama-se *hub*). Quase metade dos nodos têm grau um (ou seja,

$p_1 = 0,48$), enquanto que temos apenas um nodo com o grau maior (ou seja, $p_{92} = 1/N = 0,0005$).

Distribuições de graus observadas em vários grafos representando redes reais, sugerem uma tendência para essas redes terem distribuições seguindo uma lei de potência $p(k) \sim k^{-\gamma}$, com $2 < \gamma < 3$. Se representarmos estas redes numa escala logarítmica a sua inclinação aproxima-se duma linha recta (com declive $-\gamma$).

Em redes orientadas, temos uma distribuição de "de entrada" (in-degree) p_k^{in} e uma distribuição de graus "de saída" (out-degree) p_k^{out} .

Podemos considerar a distribuição conjunta

$$p(k, h, i) = \{\text{probabilidade } k_i^{in} = k \text{ e probabilidade } k_i^{out} = h\}$$

Se p_k^{in} e p_k^{out} são independentes então: $p(k, h, i) = p_k^{in} p_k^{out}$.

$p(k)$ tem uma "cauda" em declínio na lei de potência o que significa que a grande maioria dos nodos tem grau baixo e que existem poucos nodos, os chamados hubs, que possuem um grau elevado. Estas redes são chamadas "**livres de escala**" (**scale-free**), uma vez que aparentemente não há uma escala 'típica' para os graus em tais sistemas. Este modelo foi introduzido por Barabási e os seus colaboradores [4, 13] em 1999.

Iremos estudar diversos tipos de redes com diferentes distribuições de graus.

B.2.5 Correlação dos graus

Além de conhecermos os graus dos vértices, muitas vezes estamos interessados em saber a relação entre os nodos da rede, a correlação dos graus. Uma abordagem mais simples é considerar apenas o grau do vértice e medir a correlação de grau entre os respectivos graus de dois vértices adjacentes.

Vamos descrever "intuitivamente" como se define esta correlação.

Primeiro, considere a expressão $\sum ((x_i - \bar{x})(y_i - \bar{y}))$. Cada termo $(x_i - \bar{x})$ mede até que ponto

o valor x_i se desvia dos valores médios observados de x . Se x e y estiverem positivamente correlacionados, esperaríamos que cada produto $(x_i - \bar{x})(y_i - \bar{y})$ também fosse positivo (e certamente diferente de zero). Essencialmente, apenas estamos a calcular a média de todos esses produtos, razão pela qual dividimos a soma pelo número total de observações, n . Para compararmos realmente tais desvios, precisamos normalizar as nossas medidas, precisamos ter certeza de que os intervalos de valores que estamos comparando são "mais ou menos iguais". Basta dividir as nossas observações pelo desvio médio, ou seja, por $\frac{1}{n} \sum (x_i - \bar{x})$. Na verdade vamos considerar o desvio padrão $\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$. **Definição:** Sejam x e y duas variáveis, para as quais temos uma série de pares de observação $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. O coeficiente de correlação $r(x, y)$ entre x e y é definido por:

$$r(x, y) = \frac{\frac{1}{n} \sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

onde $\bar{x}$ é a média sobre os x'_i s : $\bar{x} = \frac{1}{n} \sum x_i$ e $\bar{y}$ é a média sobre os y'_i s : $\bar{y} = \frac{1}{n} \sum y_i$.

Definição: A distribuição conjunta de graus de pares de vértices ligados num grafo simples com N nodos, com matriz de adjacência A é definida por:

$$W(k, k'|A) = \frac{\sum_{i \neq j} \delta_{k, k_i(A)} A_{ij} \delta_{k', k_j(A)}}{\sum_{i \neq j} A_{ij}}, \forall k, k' \geq 0$$

$W(k, k'|A)$ fornece a fração de arestas na rede que conectam um vértice de grau k a um vértice de grau k' . Portanto, uma medida útil para caracterizar correlações

Definição: A razão de correlação de graus num grafo simples com N nodos, com matriz de adjacência A é definida por:

$$\Pi(k, k'|A) = \frac{\bar{k}^2(A)}{kk'} \frac{W(k, k'|A)}{p(k|A)p(k'|A)}$$

As correlações de grau capturadas por $\Pi(k, k'|A)$ podem fornecer novas informações valiosas que não estão contidas na distribuição de graus $p(k|A)$.

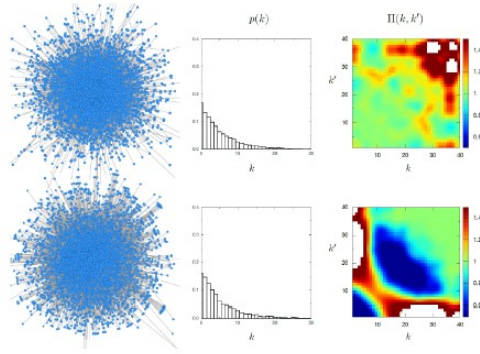


Figura B.30: Grafos G e H (à esquerda), histogramas da distribuição de graus (ao centro) e correlação de graus (à direita).

Os dois grafos simples com $N = 5000$ vértices parecem semelhantes e têm quase distribuições de graus indistinguíveis $p(k)$ (mostradas como histogramas). No entanto, eles diferem profundamente ao nível das correlações de grau, o que só é visível após o cálculo das funções $\Pi(k, k'|A)$ para os dois grafos, representados nos mapas à direita. No grafo superior, os vértices de alto grau tendem a estar mais ligados a outros vértices de alto grau. No grafo inferior, há uma forte tendência dos vértices de grau alto se conectarem com vértices de baixo grau.

B.2.6 Caminhos e conectividade

Será que dados dois vértices dum grafo existirá sempre uma aresta que os una? Ou uma sequência de vértices e arestas? Ou não?

Definição: Considere um grafo G .

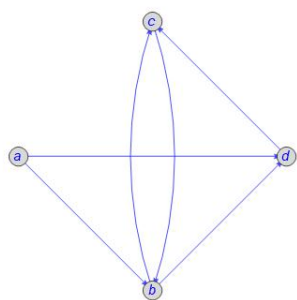
a) Um **passeio** (v_0, v_k) em G é uma sequência $\{v_0, e_1, v_1, e_2, \dots, v_{k-1}, e_k, v_k\}$ de vértices e arestas de G com $e_i = \langle v_{i-1}, v_i \rangle$. Se $v_0 = v_k$ diz-se que é um passeio fechado.

b) Um *trilho* é um passeio na qual todas as arestas são distintas e um *caminho* é um trilho no qual todos os vértices são distintos.

c) Um **ciclo** é um trilho fechado no qual todos os vértices, exceto v_0 e v_k , são distintos.

As definições para passeios, caminhos e ciclos num grafo orientado são semelhantes às definições que utilizámos para grafos, exceto que a direção das arestas deve ser consistente com a ordem em que passeio é percorrido.

Definição: O **comprimento de um caminho** representa o número de arestas que o caminho contém.



a, b, c, b, d é um passeio,
 a, b, d é um caminho,
 d, c, b, c, b, d é um passeio fechado, e
 b, d, c, b é um ciclo

Figura B.31: Grafo orientado. Exemplos de passeio, caminhos e ciclo.

Observe que b, c, b também é um ciclo para o gráfico da figura B.31. Este é um ciclo de comprimento 2. Tais ciclos não são possíveis com grafos. Observe também que c, b, a, d não é um passeio, pois $b \rightarrow a$ não é uma aresta neste grafo.

Um caminho ou ciclo num grafo orientado é dito **hamiltoniano** se visitar todos os vértices do grafo e é dito **euleriano** se contém todas as arestas. Por exemplo, a, b, d, c é o único caminho hamiltoniano para o grafo da figura anterior e não existe nenhum ciclo euleriano. Claro que esta denominação está relacionado com o famoso problemas das ponte de Königsberg identificado e resolvido por Euler em 1736.

Podemos definir os conceitos anteriores através das matrizes de adjacência do grafo. O cálculo dos parametros torna-se muito mais eficiente, nomeadamente quando consideramos grafos com muitos vértices e/ou arestas.

Consideremos produtos de entradas da matriz A de adjacência do grafo da seguinte forma, com $i, j \in \{1, 2, \dots, N\}$ e com $i_l \in \{1, 2, \dots, N\}$, para todo o l :

$$\prod_{l=1}^{k-1} A_{i_l i_{l+1}} = \overbrace{A_{i_1 i_2} A_{i_2 i_3} \dots A_{i_{k-1} i_k}}^{k-1 \text{ factores}}$$

Como cada factor individualmente é 0 ou 1, ao considerarmos os produtos, obtemos que:

$$\begin{aligned} \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 1 \text{ se } A_{i_l i_{l+1}} = 1 \forall l \in \{1, 2, \dots, k-1\} \\ \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 0, \text{ no caso contrário.} \\ \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 1 \text{ se } A_{i_l i_{l+1}} = 1 \forall l \in \{1, 2, \dots, k-1\} \\ \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 0, \text{ no caso contrário.} \end{aligned}$$

o que implica que:

$$\begin{aligned} \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 1 \text{ se o grafo contém o caminho } i_k \rightarrow i_{k-1} \rightarrow \dots i_2 \rightarrow i_1 \\ \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 0, \text{ se não contém.} \\ \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 1 \text{ se o grafo contém o caminho } i_k \rightarrow i_{k-1} \rightarrow \dots i_2 \rightarrow i_1 \\ \prod_{l=1}^{k-1} A_{i_l i_{l+1}} &= 0, \text{ se não contém.} \end{aligned}$$

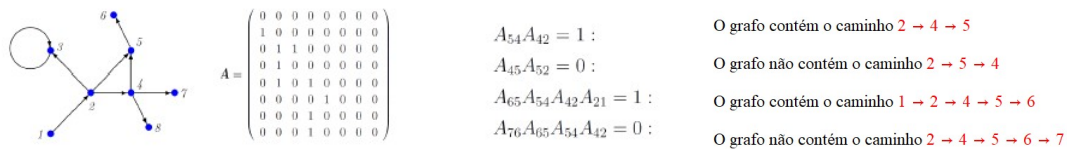


Figura B.32: Exemplo de determinação da existência de caminhos num grafo orientado através da matrix de adjacência.

Definição: Um **caminho fechado** é um caminho que começa e termina no mesmo vértice, então:

$$\prod_{l=1}^{k-1} A_{i_l i_{l+1}} = 1 \text{ com } i_1 = i_k \text{ se o grafo contém o caminho } i_k \rightarrow i_{k-1} \rightarrow \dots i_2 \rightarrow i_1$$

$$\prod_{l=1}^{k-1} A_{i_l i_{l+1}} = 0, \text{ com } i_1 = i_k, \text{ se não contém.}$$

$$\prod_{l=1}^{k-1} A_{i_l i_{l+1}} = 1 \text{ com } i_1 = i_k \text{ se o grafo contém o caminho } i_k \rightarrow i_{k-1} \rightarrow \dots i_2 \rightarrow i_1$$

$$\prod_{l=1}^{k-1} A_{i_l i_{l+1}} = 0, \text{ com } i_1 = i_k, \text{ se não contém.}$$

Definição: Um **ciclo** num grafo é um caminho que não contém vértices ou arestas repetidos, ou seja, $\prod_{l=1}^{k-1} A_{i_l i_{l+1}} = 1$ com $i_1 = i_k$ e todos os nomes dos vértices em $\{i_1, \dots, i_k\}$ não são idênticos.

Também podemos perguntar se existe algum caminho de comprimento não especificado a partir de um dado vértice inicial j para um vértice de destino especificado i . Agora não é relevante conhecer os vértices intermediários visitados entre i e j . Estamos interessados na presença ou ausência de caminhos de comprimento 2 ou maior. Mais uma vez, através da matriz de adjacência, basta considerarmos as entradas da matriz, (se não nos importamos com comprimentos de caminho mas apenas com a existência) e podemos escrever:

$$\sum_{k \geq 0} (A^{k+1})_{ij} > 0 \Leftrightarrow \text{existe pelo menos um caminho de } j \text{ para } i.$$

$$\sum_{k \geq 0} (A^{k+1})_{ij} = 0 \Leftrightarrow \text{não existe nenhum caminho de } j \text{ para } i.$$

$$\sum_{k \geq 0} (A^{k+1})_{ij} > 0 \Leftrightarrow \text{existe pelo menos um caminho de } j \text{ para } i.$$

$$\sum_{k \geq 0} (A^{k+1})_{ij} = 0 \Leftrightarrow \text{não existe nenhum caminho de } j \text{ para } i.$$

Definição: Num grafo, diz-se que os vértices i e j estão **ligados** se existir um caminho entre eles. Eles são **desligados** se tal caminho não existir.

Definição: Uma rede diz-se **conexa** se todos os pares de nodos da rede estiverem ligados. Uma rede é **desconexa** se existir pelo menos um par nodos que não está conectado.

Definição: Uma **componente** é um subconjunto de nodos numa rede, de modo que exista um caminho entre quaisquer dois nodos que pertençam a este subconjunto, mas tal que não se possa adicionar mais nodos a este subconjunto com a mesma propriedade.

Se uma rede for formada por duas componentes, e uma única aresta pode ligar estas componentes tornando a rede conexa, diz-se que esta aresta é uma ponte.

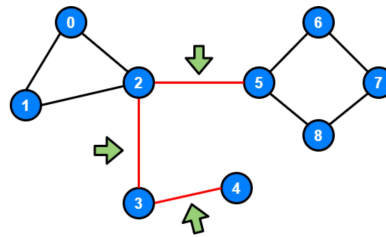


Figura B.33: Neste grafo todas as pontes estão pintadas de vermelho.

Neste grafo, na Figura B.33, é fácil identificar as pontes existentes, mas se considerarmos uma rede composta por milhões de nodos, esta identificação torna-se um grande problema, precisamos de ferramentas matemáticas, ferramentas algorítmicas que nos possam ajudar. Uma das formas é usar a matriz de adjacência e decompô-la em blocos na diagonal.

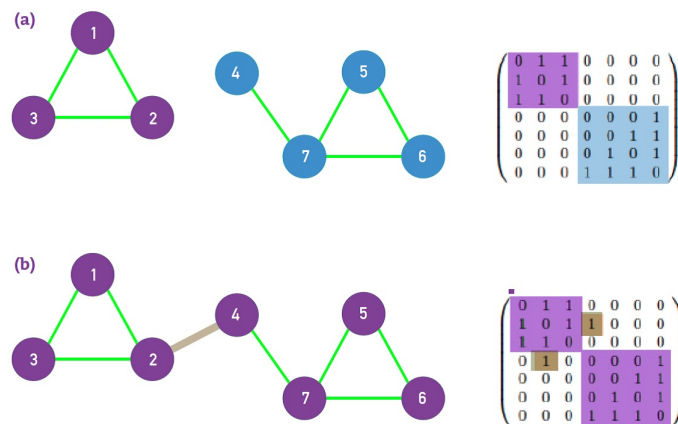


Figura B.34: No gráfico (a) temos uma rede desconexa e a respectiva matriz de adjacência. No gráfico (b) temos uma rede conexa através duma aresta , ponte, e a respectiva matriz de adjacência.

Na Figura B.34 em (a) consideramos uma rede desconexa formada por duas componentes que não estão ligadas, não há caminhos entre nodos que pertençam às diferentes componentes.

No painel direito está a matriz de adjacência da rede. Se a rede desconexa tiver componentes, a matriz de adjacência pode ser reorganizada numa forma diagonal, tal que todos os elementos diferentes de zero da matriz estão contidos em blocos quadrados ao longo da diagonal da matriz e todos os outros elementos são zero.

Na figura B.34 em (b) a adição de uma única aresta, a que chamamos ponte, colorida a cinzento, transforma uma rede desconexa numa única componente conexa, ou seja, numa rede conexa. Agora há um caminho entre cada par de nodos desta rede. Consequentemente a matriz de adjacência não pode ser escrita em forma de blocos na diagonal.

Definição: Um grafo diz-se **acíclico** se não incluir nenhum ciclo.

Definição: Quando um grafo é simples, conexo e acíclico diz-se uma **árvore**. Quando num grafo simples as únicas componentes são árvores, diz-se que é uma **floresta**. Uma **folha** num grafo árvore é qualquer vértice de grau 1.

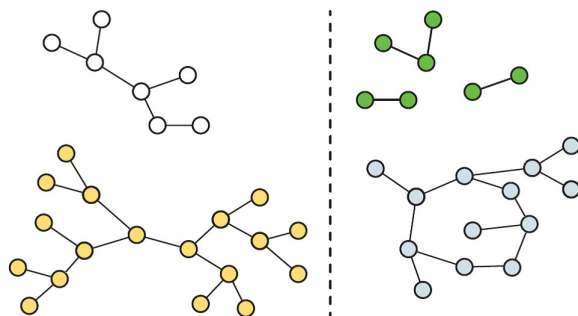


Figura B.35: Os dois grafos à esquerda (vértices brancos e amarelos) são árvores, mas os dois à direita não são: o do canto superior direito (com vértices verdes) tem vários componentes (e, portanto, não é conexo) enquanto o do canto inferior direito (vértices azuis) contém um ciclo. O gráfico no canto superior direito é, no entanto, uma floresta, pois cada uma das suas componentes é uma árvore.

Tais nomes estão relacionados com a visualização destes grafos.

Podemos afirmar o seguinte.

Se G é um grafo simples com n vértices, então são equivalentes as seguintes afirmações:

- a) G é uma árvore;
- b) G não contém ciclos e tem $n - 1$ arestas;
- c) G é conexo e tem $n - 1$ arestas;
- d) G é conexo e cada aresta é uma ponte;
- e) Quaisquer dois vértices de G estão ligados por um único caminho;
- f) G não contém ciclos, mas acrescentando uma aresta obtém-se um ciclo.

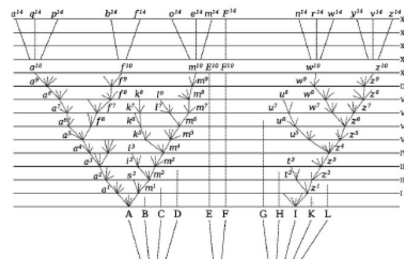


Figura B.36: A árvore evolutiva proposta por Charles Darwin (1809-1882) é uma representação hierárquica das relações evolutivas entre espécies biológicas que compartilham um ancestral comum.

Um **grafo acíclico orientado**, (em inglês: directed acyclic graph, ou simplesmente um **DAG**)

é uma rede orientada sem ciclos, por exemplo uma rede de citações ou genealogias.

B.2.7 Distância

Em redes, o que significará a distância entre vértices?

Não podemos pensar na distância "física", dois investigadores podem estar ligados porque participaram na publicação do mesmo artigo e residirem em pontos opostos do planeta, enquanto dois investigadores podem viver um ao lado do outro e não estão ligados nesta rede pois não

publicaram nenhum artigo em conjunto.

A distância entre dois pontos x, y numa variedade Riemanniana é definida como o caminho com o menor comprimento entre x e y , é conhecido como geodésica. Esta noção, *caminho com o menor comprimento*, vai ter a sua tradução na teoria de grafos.

Definição: O *caminho mais curto* (*shortest path*) entre os vértices i e j é o caminho com o menor número de arestas. O caminho mais curto é frequentemente chamado de distância entre os vértices i e j e denotado por d_{ij} .

Podemos escrever esta definição usando a matriz de adjacência.

Definição: Considere o grafo G com a matriz de adjacência A . A distância entre os vértices i e j é definida por:

A distância d_{ij} entre os vértices i e j é definida por:

- $d_{ij} = d_{ij}(A) = \infty$ entre os vértices se não existir nenhum caminho entre i e j ,
- $d_{ij} = d_{ij}(A) = l$, sendo $l \geq 0$ o mais pequeno que verifica $(A^l)_{ij} > 0$.

Podemos ter vários "caminhos mais curtos" com o mesmo comprimento entre um par de vértices.

O caminho mais curto nunca contém laços ou se cruza.

Num grafo a distância entre os vértices i e j é a mesma que a distância os vértices j e i . Num grafo orientado muitas vezes a distância não é igual. Além disso, num grafo orientado a existência de um caminho do vértice i para o vértice j não garante a existência de um caminho do vértice j para o vértice i .

Nota: observemos no caso de grafos pesados a distância entre dois vértices i e j é, frequentemente, definida em termos de um caminho com *peso* mínimo.

Definição: A *média dos comprimentos dos caminho mais curtos* (average shortest path length) num grafo, com N vértices, com matriz de adjacência A é:

$$\text{grafos: } \bar{d}(A) = \frac{2}{N(N-1)} \sum_{i < j} d_{ij}(A).$$

$$\text{grafos orientados: } \bar{d}(A) = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}(A)$$

Definição: O *diâmetro* de um grafo G , com matriz de adjacência A é definido como $d(A) = \max_{i \neq j} dij(A)$ (ou seja, a distância entre o par de vértices que estão mais distantes um do outro no grafo).

Definição: Consideremos um vértice $v_1 \in V$. A *excentricidade* de v_1 é a maior distância de v_1 a qualquer outro vértice $v_2 \in V$: $ecc(v_1) = \max_{v_2 \in V} d(v_1, v_2)$.

Definição: O *raio* de G é a excentricidade mínima de qualquer vértice em V : $rad(G) = \min_{v_1 \in V} ecc(v_1) = \min_{v_1 \in V} \max_{v_2 \in V} d(v_1, v_2)$.

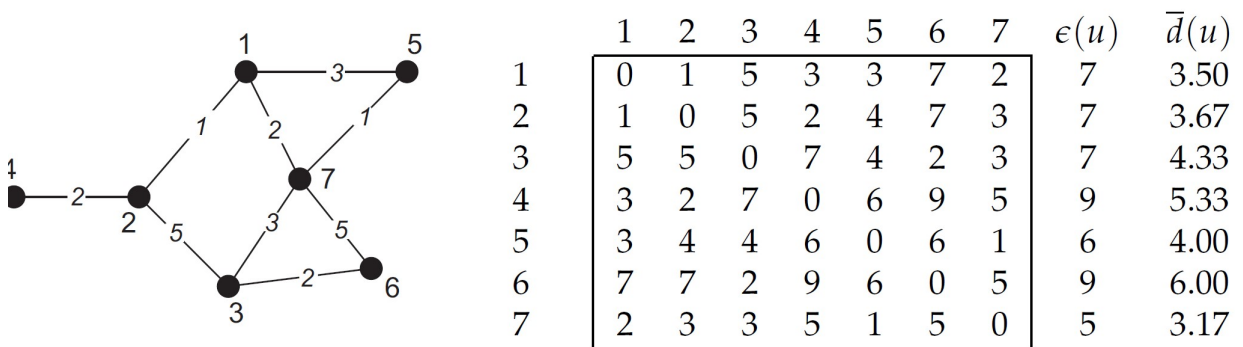


Figura B.37: À esquerda o grafo pesado W , ao centro a sua matriz de adjacência e no lado direito as coluna com as excentricidades de cada um dos vértice e os comprimento médios.

Na figura B.37 está desenhado um grafo pesado W , a matriz com os caminhos entre os vértices e coluna com as excentricidades de cada vértice. Podemos calcular que o raio do grafo é igual a 5

e que o diâmetro é igual a 9.

Definição: Se existir um ciclo em G , então o **perímetro** (girth) de G é o comprimento do ciclo mais curto. Quando não existe nenhum ciclo em G diz-se que o perímetro é zero.

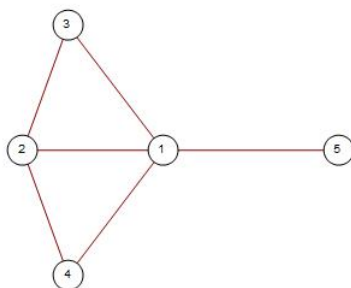


Figura B.38: Neste grafo o diâmetro é 2, o raio é 1 e o perímetro é 3.

Esta medida é muito importante no estudo da análise espectral dos grafos.

B.3 Agrupamento - clustering

Um aspecto importante da rede é o comportamento de agrupamento (clustering), tanto numa escala local como numa escala global. Do ponto de vista local o coeficiente de agrupamento diz-nos a probabilidade média de dois vizinhos de algum vértice serem diretamente ligados. O ponto de vista global, também é chamado estrutura da comunidade. Neste caso procura-se os grupos de vértices com muitas conexões dentro de um grupo, mas consideravelmente menos entre os grupos.

Consideremos em primeiro lugar a escala local.

Duma forma simples, queremos ver, para um dado vértice, até que ponto os vizinhos deste vértice, também são vizinhos uns dos outros. O coeficiente de agrupamento c_i de um vértice v_i é dado pela proporção de ligações entre os vértices dentro de sua vizinhança dividida pelo número de ligações que podem existir entre eles.

Definição: Denotando por k_i o grau do vértice v_i e por $N_i = \{v_j : e_{ij} \in E\}$ o conjunto dos

vizinhos de v_i , então para um grafo, podemos definir, o coeficiente de agrupamento c_i do vértice v_i e a média dos coeficiente de agrupamento c do grafo por:

$$c_i = \frac{2|\{e_{ij}\}|}{k_i(k_i - 1)}, \text{ com } e_{ij} \in E; v_j, v_k \in N_i \text{ e } c = \frac{1}{N} \sum_{i=1}^N c_i.$$

Nos casos extremos temos que $c_i = 0$ se nenhum dos vizinhos do vértice v_i se conectam e $c_i = 1$ se os vizinhos do vértice v_i formam um grafo completo, ou seja, todos eles ligam entre si.

Numa linguagem probabilística, c_i é a probabilidade de que dois vizinhos de um vértice se conectem. Por exemplo se, $c_i = 0,5$ tal valor implica que há 50% de hipóteses de que dois vizinhos de um vértice estejam ligados.

Podemos definir o coeficiente de agrupamento, utilizando a matriz de adjacência.

Definição: O coeficiente de agrupamento $c_i(A)$ do vértice i com grau ≥ 2 num grafo com N vértices e com matriz de adjacência A é definido por:

$$\begin{aligned} c_i &= \frac{\text{número de pares de vértices ligados entre vizinhos de } i}{k_i(k_i - 1)} = \\ &= \frac{\sum_{j,k=1}^N (1 - \delta_{jk}) A_{ij} A_{jk} A_{ik}}{\sum_{j,k=1}^N (1 - \delta_{jk}) A_{ij} A_{ik}} \end{aligned}$$

Como os produtos de entradas da matriz de adjacência de um grafo podem ser usados para identificar caminhos, podemos contar o número de caminhos fechados de um determinado comprimento, usando a matriz de adjacência.

O número L_l de caminhos fechados de comprimento $l > 0$ num grafo com N vértices e com matriz de adjacência A é dado por

$$L_l = \sum_{i_1=1}^N \dots \sum_{i_l=1}^N \left(\prod_{k=1}^{l-1} A_{i_k i_{k+1}} \right) A_{i_l, i_1} = \sum_{i=1}^N (A^l)_{ii} = \text{Tr}(A^l).$$

Definição: O número de triângulos T_i envolvendo o vértice i num grafo simples com N vértices e com matriz de adjacência A é definido como $T_i = \frac{1}{2} \sum_{j,k=1}^N A_{ij} A_{jk} A_{ik} \in \mathbb{N}$.

T_i , que também pode ser escrito como $T_i = \frac{1}{2} (A^3)_{ii}$, conta o número de caminhos fechados de comprimento três, nos quais o vértice i participa. Observe que em grafos simples tem-se $c_i = \frac{2T_i}{k_i(k_i - 1)}$.

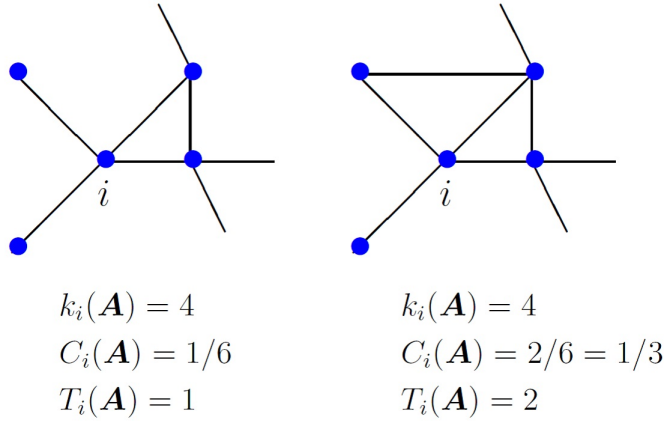


Figura B.39: Nos dois grafos temos os graus do vértice, o número de triângulos e o coeficiente de agrupamento.

Consideremos agora o ponto de vista global. Tentamos encontrar métodos que nos permitam detectar os tais subgrupos, que denominamos por comunidades.

Significa que tentamos encontrar subgrafos, os clusters, que são internamente densamente ligados (dentro), mas escassamente ligados com o resto do grafo. Apenas exemplificamos um destes métodos pois este método é um importante exemplo da íntima ligação da teoria das redes com a geometria diferencial. Este método é baseado numa quantidade, $h(V)$, que é análoga à **constante**

de Cheeger, introduzida por Cheeger na geometria Riemanniana.

Seja k_i o grau do vértice i , $|E|$ o número de arestas do conjunto de arestas E e E_0 é o subconjunto de arestas que se forem removidas transformam as componentes V_1, V_2 desconexas.

$$h(V) = \inf \left[\frac{|E_0|}{\min \left(\sum_{i \in V_1} k_i, \sum_{i \in (V \setminus V_1)} k_i \right)} \right]$$

Assim, tentamos dividir o grafo em duas grandes componentes removendo apenas algumas arestas. Podemos repetir este processo em cada uma destas componentes até que não seja possível calcular o valor da quantidade h

Na verdade, tentamos desconectar o grafo removendo o menor número de arestas possível, para obter dois "grandes clusters" e, iteramos este processo até ser possível.

B.3.1 Parte espectral

A teoria de redes dinâmicas está preocupada com sistemas de unidades acopladas de acordo com uma estrutura subjacente do grafo. Por isso, investiga a interação entre dinâmica e estrutura, entre os processos temporais acontecendo nas unidades individuais e a estrutura espacial estática que as liga.

Para analisar essa estrutura espacial, formalizada como um grafo, discutimos um sistema completo de invariantes do grafo, o espectro do Laplaciano do grafo, e como se relaciona com várias propriedades qualitativas do grafo.

Em seguida, voltamos-nos para aspectos dinâmicos e discutimos sistemas de osciladores com acoplamento de acordo com o Laplaciano do grafo e analisamos a sua sincronizabilidade.

A ferramenta analítica aqui são expansões locais em termos de vectores próprios do Laplaciano

do grafo. Isto é visto como um primeiro passo para uma compreensão geral do padrão da formação em sistemas de osciladores acoplados.

O objectivo da teoria espectral dos grafos é entender as propriedades dos grafos, usando ferramentas de álgebra linear, particularmente usando valores próprios e vectores próprios de várias matrizes associadas a grafos, no nosso caso estudaremos sobretudo o caso da matriz laplaciana.

Assim como os astrónomos estudam espectros estelares para determinar a composição de estrelas, um dos principais objectivos da teoria dos grafos é deduzir as principais propriedades e estrutura de um grafo a partir de seu espectro (ou de uma pequena lista de invariantes computáveis). A abordagem espectral para grafos gerais é um passo nesta direção. Veremos que os valores estão intimamente relacionados com quase todos os principais invariantes de um grafo.

Já definimos a matriz de adjacência para grafos simples, orientados e pesados.

Consideremos um grafo G com N vértices. Definimos a matriz **diagonal** dos graus D como sendo a matriz diagonal $N \times N$ dada por $D = D(G) = [d_{ij}]$, where $d_{ii} = k_i$.

Definição: A matriz **Laplaciana** L de um grafo com N nodos, com matriz de adjacência A , é definida por:

$$L = D - A$$

Esta matriz tem algumas características interessantes e propriedades dinâmicas muito úteis: pelo menos um de seus valores próprios é zero e todos os outros valores próprios são zero ou positivos e estão contidos no intervalo $[0, \min \{N, 2\Delta\}]$, onde Δ é o maior grau dos vértices.

O espectro de L pode ser ordenado, $\lambda_1 = 0 \leq \lambda_2 \leq \dots \leq \lambda_N$. O segundo valor próprio λ_2 é conhecido como conectividade algébrica ou valor de Fiedler e desempenha um papel especial,

veremos que determina a rapidez com que a difusão ocorre na rede.

Prova-se, ainda, que a multiplicidade do valor próprio nulo λ_1 é igual ao número de componentes conexas do grafo.

Muitas vezes é mais conveniente estudar a matriz laplaciana normalizada.

Definição: A matriz **Laplaciana normalizada** $\mathcal{L}$ de um grafo simples com N nodos, é definida por

$$\mathcal{L} = D^{\frac{1}{2}} L D^{-\frac{1}{2}}$$

Se definirmos em termos matriciais

Definição: A matriz **Laplaciana normalizada de $\mathbf{G}$** é a matriz $\mathcal{L} = [l_{ij}]$ onde

$$l_{ij} = \begin{cases} 1, & \text{se } i = j \text{ e } d(v_i) \neq 0; \\ -\frac{1}{\sqrt{d_i d_j}}, & \text{se } i \text{ é adjacente a } j; \\ 0, & \text{se } i \text{ não é adjacente a } j. \end{cases}$$

Usar o Laplaciano normalizado é muitas vezes conveniente, pois todos os seus valores próprios pertencem ao intervalo $[0, 2]$, o que facilita a comparação dos espectros de dois grafos. No entanto, mesmo grafos não isomórficos podem compartilhar o mesmo espectro.

Definição : Os grafos que possuem o mesmo espectro são chamados **coespectrais** (também isoespectrais). Quando a igualdade dos conjuntos de valores próprios distintos aos pares é válida, mas as multiplicidades não coincidem, os grafos são chamados **fracamente coespectrais**.

B.4 Centralidades

Uma pergunta frequente quando se estuda uma rede: quais são os vértices mais importantes da rede?

A ordem alfabética tradicional da Enciclopédia é hoje substituída na vida quotidiana por algoritmos de pesquisa na web que classificam as páginas da web (ou seja, os nodos da World Wide Web) por ordem decrescente de relevância para a consulta. A centralidade de um vértice pode refletir a importância do vértice em manter uma rede ligada, em diminuir a distância mais curta entre vértices, ou pode ser apenas uma propriedade do vértice como o seu grau.

B.4.1 Medidas mais comuns

Centralidade do grau

A centralidade de grau não é mais do que o valor cde uma grau do vértice que já definimos anteriormente.

B.4.2 Centralidade de intermediação (“Betweenness centrality”)

A centralidade de intermediação, foi proposta por Freeman e está relacionada com o número de vezes que um vértice precisa de outro vértice (cuja centralidade é medida) com o objectivo de alcançar um terceiro vértice. Mede o papel de intermediário num grafo e dá uma ideia do volume de tráfego/informação que flui entre quaisquer dois vértices através do intermediário.

Definição: *Seja G um grafo (conexo ou não) com n vértices e seja v_k um vértice de G . Considere um par de vértices v_i e v_j em G , tal que $i \neq j$, $i \neq k$ e $j \neq k$. A intermediação parcial de v_k com respeito a v_i e v_j é dada por:*

$$b_{ij}(v_k) = \begin{cases} 0, & \text{se não existir caminho entre } v_i \text{ e } v_j; \\ \frac{g_{ij}(v_k)}{g_{ij}}, & \text{caso contrário,} \end{cases}$$

onde g_{ij} denota o número de geodésicas entre v_i e v_j e $g_{ij}(v_k)$ denota o número de geodésicas entre v_i e v_j que passam por v_k .

Para se determinar a centralidade de intermediação de um vértice v_k num grafo G com n vértices, basta somar todas as intermediações parciais de v_k em G .

Definição: Nas condições da definição anterior a **centralidade de intermediação** de um vértice v_k num grafo G com n vértices, denotada por $c_B(v_k)$ é dada por:

$$c_B(v_k) = \sum_{\substack{1 \leq i \leq j \leq n \\ i, j \neq k}} b_{ij}(v_k).$$

Como é visível no exemplo simples da Figura B.40 a centralidade de intermediação do vértice 8 é elevada.

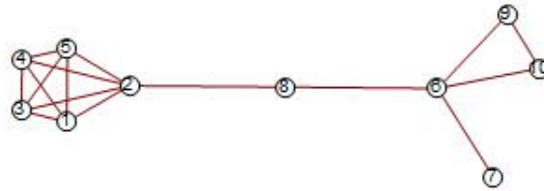


Figura B.40: $Cb(8) = 10^2 + 5^2 - 4^2 = 59$.

B.4.3 Centralidade de Proximidade “Closeness centrality”)

A ideia principal por trás da definição de centralidade de proximidade é classificar a importância de um vértice em função da sua distância aos restantes vértices da rede. Portanto a centralidade de proximidade é menor para vértices que têm uma maior distância média aos restantes vértices da rede.

Definição: Seja G um grafo conexo com n vértices e seja v_k um vértice de G . A centralidade de proximidade de v_k é dada pelo inverso da soma das distâncias de v_k a todos os outros vértices do grafo, ou seja,

$$C_C(v_k) = \frac{1}{\sum_{1 \leq j \leq n} d(v_j, v_k)}.$$

Centralidade de eficiência

Os problemas de determinação dum local de modo que minimize o tempo máximo de viagem entre o mesmo e todas as demais localizações é um assunto muito estudado, por exemplo, a instalação

de um hospital, cujo objectivo é minimizar o tempo máximo de atendimento de uma ambulância a uma possível emergência. Neste sentido foi proposta uma medida chamada centralidade eficiência, baseada no conceito de excentricidade de um vértice.

Definição: Seja G um grafo conexo com n vértices e seja v_k um vértice de G . A centralidade de eficiência de v_k é dada pelo inverso da excentricidade de v_k , isto é,

$$c_{efi}(v_k) = \frac{1}{ecc(v_k)},$$

$$\text{onde } ecc(v_k) = \max_{v_i \in N(v_k)} d(v_i, v_k).$$

B.4.4 Medidas de centralidade espectrais

Estudo de medidas de centralidade tendo como suporte a caracterização espectral das redes

Centralidade do vector próprio

A centralidade de vector próprio é uma medida proposta por Bonacich em 1987 (??) que corresponde à relevância de um vértice em função da relação com os seus vizinhos. *O importante não é quantas pessoas se conhece, mas quem se conhece.*

Definição: Seja G um grafo com n vértices, x um vértice de G , a_{jx} elementos da matriz de adjacência A e seja $\lambda_{\max(A)}$ o maior valor próprio (em módulo) da matriz de adjacência e $v = (v_1, \dots, v_n)^T$ o vector próprio associado. Definimos a **centralidade do vector próprio** por:

$$\sigma_E(x) = \sum_{j=1}^n a_{jx} \cdot v_j.$$

Mas numa rede orientada em que não existe nenhuma componente fortemente ligada todos os vértices têm centralidade nula. Para tentar resolver esta debilidade é proposta uma outra medida.

Centralidade de Katz

Para resolver esta debilidade, ou seja, o desaparecimento da centralidade do vector próprio para os vértices na componente interna de rede orientada, foi proposta a centralidade Katz. A centralidade de Katz atribui a cada vértice um pequeno valor β apenas por ser vértice da rede. Logo a centralidade do vértice só aumenta se muitos vértices importantes apontarem para este vértice.

Definição: A **centralidade de Katz** x satisfaz a seguinte equação.

$$x_i = \alpha \sum_{j=1}^n a_{ij} x_j + \beta.$$

onde $\beta > 0$ e $\alpha \in \left(0, \frac{1}{\lambda_1}\right)$ onde λ_1 é o valor próprio de the Perron-Frobenious da matriz de adjacência A .

Usando o formalismo matricial podemos escrever esta equação como

$$x = \alpha Ax + \beta \mathbf{1}$$

em que $\mathbf{1}$ indica o vector coluna tal que $\mathbf{1}_i = 1, \forall i = 1, \dots, n$.

Com cálculo algébrico podemos ainda escrever

$$x = (I - \alpha A)^{-1} + \beta \mathbf{1} = \beta \sum_{n=0}^{\infty} (\alpha A)^n \mathbf{1}$$

A matriz $(I - \alpha A)^{-1}$ diverge para $\det(I - \alpha A) = 0$.

Assim, para garantir a convergência $(I - \alpha A)^{-1}$ e uma centralidade Katz bem definida que devemos considerar o parâmetro $\alpha \in \left(0, \frac{1}{\lambda_1}\right)$.

Centralidade PageRank

A centralidade do PageRank foi o principal algoritmo por trás do motor de busca Google, e desempenhou um papel fundamental na determinação do sucesso da Google.

A centralidade Katz da secção anterior tem uma característica que pode ser indesejável. Se um vértice com elevada centralidade de Katz aponta para muitos outros, então esses vértices também obtêm elevada centralidade. Um vértice de elevada centralidade apontando para um milhão de outros conferem a um milhão deles elevada centralidade. O aumento da centralidade em virtude de receber uma aresta de um vértice de prestígio é diluída por ser partilhada com tantos outros. Por exemplo, o famoso Yahoo! pode conter uma ligação para a minha página web, mas também contém ligações para milhões de outras páginas.

Os estudantes de doutoramento da Universidade de Stanford, Larry Page e Sergey Brin, os fundadores do PageRank, publicam em Abril de 1998 o artigo “The anatomy of a large-scale hypertextual Web search engine”, no dia 1 de Setembro de 1998 registam a primeira patente do Page Rank e no dia 4 de Setembro de 1998 a Google é incorporada.

Ser Google não é fácil - três problemas principais na recuperação de informação: sinonímia (múltiplas formas de dizer a mesma coisa); polissemia (múltiplos significados para o mesmo termo) e quantidade (existem imensas páginas web, e a maioria é irrelevante para um utilizador do motor de pesquisa).

A premissa do Page Rank é que um hiperlink de uma página para outra é um voto para a página de destino.

Para resolver a questão que descrevemos, no PageRank os vértices a jusante adquirem apenas uma fracção da centralidade do próprio vértice central que aponta para eles. Definimos uma variação da centralidade de Katz em que a centralidade que obtenho dos meus vizinhos da rede



Figura B.41: PageRank

é proporcional à sua centralidade dividido pelo seu grau de saída. Assim, vértices que apontam para muitos outros passam apenas uma pequena quantidade de centralidade para cada um desses vértices, mesmo que a sua centralidade seja elevada.

Definição: A **centralidade de PageRank** x satisfaz a seguinte equação.

$$x_i = \alpha \sum_{j=1}^n a_{ij} \frac{1}{\kappa_j^{out}} x_j + \beta.$$

onde $\kappa_j^{out} = \max(k_j^{out}, 1)$ $\beta > 0$ e $\alpha \in \left(0, \frac{1}{\lambda_1}\right)$ onde λ_1 é o valor próprio de the Perron-Frobenius da matriz AD^{-1} com D é a matriz diagonal definida pelos elementos $D_{ii} = \kappa_i^{out} = \max(k_i^{out}, 1)$.

Usando o formalismo matricial podemos escrever esta equação como

$$x = \alpha AD^{-1}x + \beta \mathbf{1}$$

em que $\mathbf{1}$ indica o vector coluna tal que $\mathbf{1}_i = 1, \forall i = 1, \dots, n$.

Com cálculo algébrico podemos ainda escrever

$$x = (I - \alpha AD^{-1})^{-1} \beta \mathbf{1} = \beta \sum_{n=0}^{\infty} (\alpha AD^{-1})^n \mathbf{1}$$

A condição $\alpha < \frac{1}{\lambda_1}$ garante que a centralidade PageRank está bem definida para todos os vértices da rede. Numa rede simples $\lambda_1 = 1$ e nesse caso $\alpha \in (0, 1)$. No algoritmo original PageRank da Google $\alpha = 0.85$.

Comparação de centralidades

Vamos comparar alguma medida de centralidade para a mesma rede supondo que os valores das centralidades variam entre o vermelho, que é o mais alto, e o azul é o valor de centralidade mais baixo.

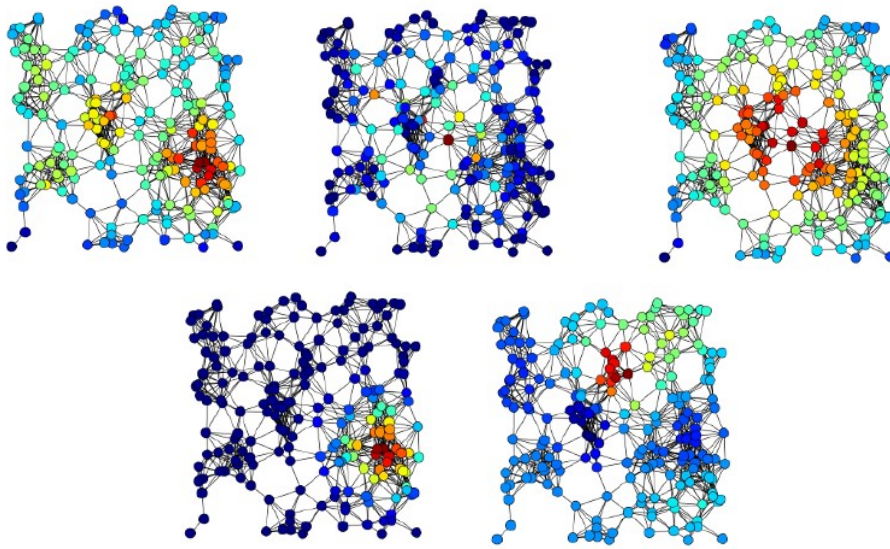


Figura B.42: Na parte superior temos no lado esquerdo a centralidade de grau, ao centro a centralidade de intermediação e no lado direito a centralidade de proximidade. Na parte inferior temos no lado esquerdo a centralidade do vector próprio e no lado direito a centralidade de Katz.

Podemos observar na Figura B.42, por exemplo, que se compararmos a centralidade de grau e a centralidade do vector próprio, na centralidade de vector próprio existe uma maior distinção entre vértices com elevada centralidade e nodos com baixa centralidade.