

Entre os dados qualitativos e a análise de redes

Albertina Ferreira¹ Carlos Caldeira² Fernanda Olival³

¹ Instituto Politécnico de Santarém; albertina.ferreira@esa.ipsantarem.pt

² Universidade de Évora; ccaldeira@di.uevora.pt

³ Universidade de Évora; mfo@uevora.pt

Resumo A preparação de dados de modo a adequá-los para serem utilizados em análise de redes é um aspeto fundamental nas bases de dados prosopográficas que envolvem o registo de relações. Mais complexo é ainda quando esses dados se reportam a sociedades pretéritas (séculos XVI a XVIII), como é o caso.

Neste ensaio referimos a importância da identificação de ocorrências anómalas, resultantes da introdução incorreta de dados, e mencionamos ainda as metodologias seguidas para identificar algumas dessas situações. Sugerimos também os procedimentos a seguir para colocar os dados num formato adequado à sua integração em *software* de análise de redes. Por último, apresentamos alguns dos resultados obtidos para a rede que analisámos.

O repositório de dados que utilizámos tem armazenada informação sobre eventos biográficos e relacionais, sendo o tratamento dos dados fundamental para o estudo das redes de relações entre os diversos atores sociais.

1 Introdução

O estudo da teoria de redes nas ciências físicas e sociais tem sido uma área pela qual os investigadores demonstram crescente interesse. Borgatti *et al.* [3] salientam que a teoria das redes tem possibilitado explicações para os mais diversos fenómenos sociais numa ampla diversidade de contextos.

Como atua uma rede? Como evolui? Podemos encontrar leis e derivar modelos que expliquem essa evolução [6]?

Estas questões são só um exemplo entre muitas outras que se podem colocar. Para que consigamos obter respostas para todas estas questões, há que percorrer um caminho: desde a obtenção de dados, passando pela sua manipulação e exploração, culminando com a análise efetiva da rede.

Embora os autores anteriormente citados considerem essencialmente redes a funcionar na atualidade, grande parte dos estudos que realizam poderão ser estendidos a outras épocas, bem como a outras sociedades. Neste trabalho abordamos o tratamento dos dados e referimos de que modo este se torna importante na análise dos diversos relacionamentos, entre os vários intervenientes nos processos de Familiaturas do Santo Ofício (SO). Este estudo enquadra-se numa das tarefas propostas - *Developing SPAREs*⁴: *social network analysis* - do projeto aprovado e financiado pela FCT⁵: PTDC/HIS-HIS/118227/2010 – CIDEHUS⁶.

Os objetivos deste trabalho são: 1) analisar a importância da preparação dos dados qualitativos para serem tratados pelo *software* disponível para a análise de redes; além do mais há sempre erros, quando se trabalha com grandes números; 2) discutir a questão da escolha do *software* e da validação dos resultados numa equipa multidisciplinar.

Note-se que o facto de trabalharmos com dados reais e com uma equipa multidisciplinar é muito relevante, pois é sempre possível tentar garantir empiricamente o controlo dos resultados e até confrontar diferentes interpretações dos mesmos.

⁴ Sistema Prosopográfico de Análise de Relações e Eventos Sociais

⁵ Fundação para a Ciência e a Tecnologia

⁶ Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora

2 Procedimentos Metodológicos

Os dados a utilizar encontram-se distribuídos por três séculos, recaindo o nosso estudo sobre um número bastante elevado de registos (cerca de 100000), os quais se encontram disponíveis na base de dados prosopográfica *SPARES*, desenvolvida no âmbito do projeto FCOMP-01-0124-FEDER-007360 – Inquirir da Honra: Comissários do Santo Ofício e das Ordens Militares em Portugal (1570 – 1773).

Numa primeira fase automatizámos a extração dos dados a partir da base de dados prosopográfica *SPARES*, de modo a que estes possam ser manipulados por dois softwares de rede: *PAJEK* e *GEPHI*.

Inicialmente escolhemos o *PAJEK* para construirmos e analisarmos a nossa rede. Com efeito, esta aplicação consegue, por um lado, explorar e manipular redes de grande dimensão e, por outro, encontrar-se disponível gratuitamente, para uso não comercial. Pode ser acedida a partir de: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> [2] [5]. Embora através desta aplicação se consiga analisar redes e obter dados, tanto analíticos como gráficos, que podem ser explorados por outras aplicações, pesquisámos outras ferramentas *Open Source*. Como futuramente pretendemos integrar na mesma plataforma o pré-processamento dos dados e a análise de rede, consideramos que o *GEPHI* (disponível em: <http://gephi.org/>) [1] permitirá uma maior interoperabilidade. O ponto de partida para a extração dos dados é a tabela que se pode observar na Figura 1.



Nome de P1	Código de P1	Relação	Código de P2	Nome de P2
Luís de Araújo de Barros	9199	Consulta favorável no CG	6073	Marcos Teixeira
Martim Afonso de Melo	6155	Consulta favorável no CG	6073	Marcos Teixeira
Salvador de Mesquita	6135	Consulta favorável no CG	6073	Marcos Teixeira
Luís de Araújo de Barros	9199	Consulta favorável no CG	6123	António de Matos de Noronha [Dom]
Luís de Araújo de Barros	9199	Consulta favorável no CG	6088	Diogo de Sousa [Doutor]
Salvador de Mesquita	6135	Consulta favorável no CG	6088	Diogo de Sousa [Doutor]
Salvador de Mesquita	6135	Consulta favorável no CG	6123	António de Matos de Noronha [Dom]

Figura 1. Dados da base de dados *SPARES*.

Para preparar os dados de modo a poderem ser utilizados no *software* de rede foi necessário:

- criar uma tabela com os códigos e nomes de P1;
- acrescentar a essa tabela os códigos e nomes de P2;
- criar uma tabela com os vértices da rede;
- criar uma tabela com as relações da rede;
- gerar as listagens que vão ser exportadas.

Para que o ficheiro obtido pudesse ter o formato que o *PAJEK* lê foi ainda necessário:

- criar um procedimento e uma pesquisa para atribuir uma numeração sequencial;
- criar um procedimento para eliminar linhas em branco do ficheiro de *output*.

Para que se possa vir a analisar a rede por intervalo de tempo foi também necessário preparar os ficheiros com a informação da década a que cada uma das relações corresponde. A data está num formato de texto que varia entre o padrão 1709=11=08 (exatamente esta data) ou 1742<06<09 (pensa-se que tenha ocorrido antes desta data) e vagamente nesta data (1709-00-00). Os historiadores precisam de trabalhar deste modo, pois nem sempre têm a certeza da cronologia exata da ocorrência. Foi necessário alocar a cada vértice (indivíduo) as décadas em que ele interveio nos processos de familiaturas do Santo Ofício. O intervalo de tempo resultante para cada vértice deverá ter o formato [década x-década y] ou por exemplo [2-4], que significa que o comissário interveio durante 30 anos, ou seja, 3 décadas (2, 3 e 4).

No decorrer deste trabalho identificaram-se situações anómalas, as quais foram corrigidas, nomeadamente:

- datas negativas;
- datas inferiores a 1579 (primeira relação conhecida);
- comissários que mantinham relação com eles próprios;
- o mesmo código (único para cada um dos indivíduos) atribuído a dois indivíduos diferentes;
- o mesmo indivíduo com nomes diferentes.

A identificação destas ocorrências foi feita através de pesquisas quando se identificou que o ficheiro final possuía mais relações do que as originais. Tivemos de rever todo o percurso e identificámos as situações atrás referidas. Relativamente às 4 primeiras situações, foram corrigidas manualmente, pois é necessário conhecer a base de dados, nomeadamente as relações envolvidas. No que diz respeito à última relação, criou-se um procedimento que atribuisse ao nome do indivíduo referenciado em P1 o nome que ele possuía em P2.

Após obtermos os ficheiros que servirão de *input* aos *softwares* de análise de redes, realizaram-se algumas análises que permitem assim testar todo o trabalho que foi feito ao nível deste tratamento de dados.

3 Resultados e Discussão

Apresentamos na Figura 2 e na Figura 3 exemplos de ficheiros construídos para utilização nos softwares de análise de redes.

```
*Vertices 12
1 "António Correia Bethencourt" ic Red bc Red" [13-15]
2 "António de Noronha e Menezes [Dom]" ic Red bc Red" [15-18]
3 "António Mouzinho [Doutor]" ic Blue bc Blue" [1]
4 "Bartolomeu César de Andrade" ic Red bc Red" [11-15]
5 "Bento Pais do Amara" ic Green bc Green" [16-19]
6 "Cristóvão de Sousa e Lira [Licenciado]" ic Red bc Red" [13-14]
7 "Diogo Fernandes Branco" ic Red bc Red" [8-12]
8 "Jacome Esteves Nogueira" ic Blue bc Blue" [14-18]
9 "José de Sousa Castelo Branco [Dom]" ic Blue bc Blue" [12-14]
10 "Mariana Isabel de Mesquita e Noronha [Dona]" ic Red bc Red" [18]
11 "Martim Filter" ic Blue bc Blue" [8]
12 "Mateus da Silva" ic Red bc Red" [1]

*arcs
9 1 1 [13]
5 2 1 [18]
9 4 1 [14]
8 5 1 [18]
9 6 1 [13]
11 7 1 [8]
5 10 1 [18]
3 12 1 [1]
```

Figura 2. *Input* para PAJEK.

Vértices	Arestas
Id;Label	Source;Target;Label;Weight
91;Bento Pais do Amara	91;2351;Patrocínio;1.0
151;Bartolomeu César de Andrade	91;2035;Patrocínio;1.0
152;Cristóvão de Sousa e Lira [Licenciado]	153;154;Patrocínio;1.0
153;José de Sousa Castelo Branco [Dom]	153;152;Patrocínio;1.0
154;António Correia Bethencourt	153;151;Patrocínio;1.0
2035;António de Noronha e Menezes [Dom]	2375;91;Patrocínio;1.0
2351;Mariana Isabel de Mesquita e Noronha [Dona]	4731;2613;Patrocínio;1.0
2375;Jacome Esteves Nogueira	6094;6076;Patrocínio;1.0
2613;Diogo Fernandes Branco	
4731;Martim Filter	
6076;Mateus da Silva	
6094;António Mouzinho [Doutor]	

Figura 3. *Inputs* para GEPHI.

Dos cerca de 500 eventos relacionais identificados na base de dados *SPARES*, vamos analisar os comissários do concelho de Arraiolos, no que diz respeito às relações de: “Ouvida como testemunha na extra-judicial SO pelo comissário ad hoc com rol do pároco”; “Ouvida como testemunha na habilitação SO pelo comissário SO”.

Escolhemos este concelho por ser de pequenas dimensões e estes eventos por se encontrarem já estudados do ponto de vista estatístico para esta localidade. Assim, havia todo o interesse em confirmar graficamente o que analiticamente já se conhecia. Por outro lado, as relações “Ouvida como testemunha ...” fazem parte daquelas que têm um papel preponderante na cotação dos atores sociais.

Relativamente ao primeiro evento (Figura 4), podemos verificar que os dois comissários que não são de Arraiolos, e que ali atuaram entre 1743 e 1746, solicitaram ao pároco uma lista das testemunhas que deviam auscultar. Estas listas não se cruzam entre si, até porque se referiam a freguesias diferentes (S. Gregório e Igrejinha).

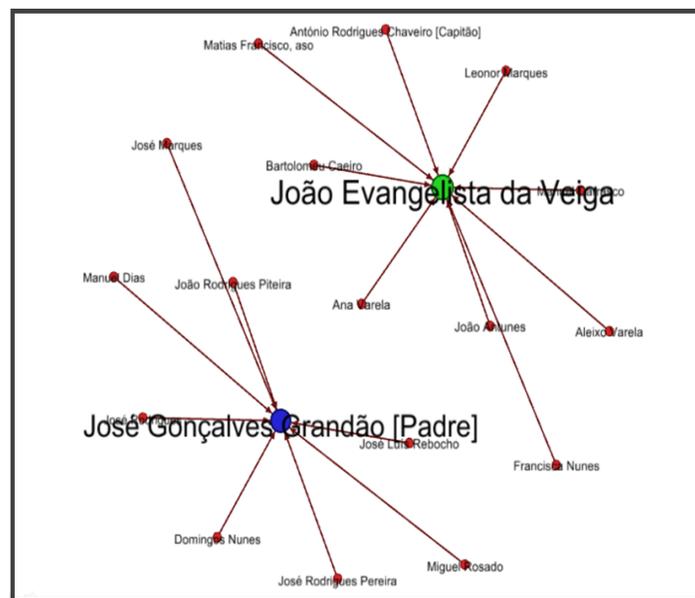


Figura 4. Evento relacional “Ouvida como testemunha na extra-judicial SO pelo comissário ad hoc com rol do pároco”.

A partir desta análise, podemos observar se as testemunhas fornecidas são das que mais intervêm nos concelhos de Arraiolos. Fez-se assim uma nova análise que veio confirmar que efetivamente estas são das testemunhas mais vezes chamadas, apesar de uma delas (Leonor Marques) ser mulher – eram menos requeridas – e não saber ler nem escrever (Figura 5).

No que diz respeito ao segundo evento, observamos que existe um maior número de intervenientes, dos quais o comissário Gaspar Barreto de Ladim é o que tem um maior número de relações (55). Este resultado veio confirmar a sua importância. Tratou-se de um comissário ativo durante 30 anos nesta rede. Na Figura 6 apresenta-se a rede, realçando os comissários mais ativos nos concelhos de Arraiolos.

Atribuímos a cor amarela àqueles que moravam em Arraiolos. O *GEPHI* permite facilmente efetuar este tipo de manipulações, de modo a realçar pormenores que o investigador pretende destacar. Neste caso, era relevante destrinçar os ali moradores (conheciam toda a gente) dos que não moravam no concelho, pois tinham menos probabilidades de conhecerem minuciosamente todos.

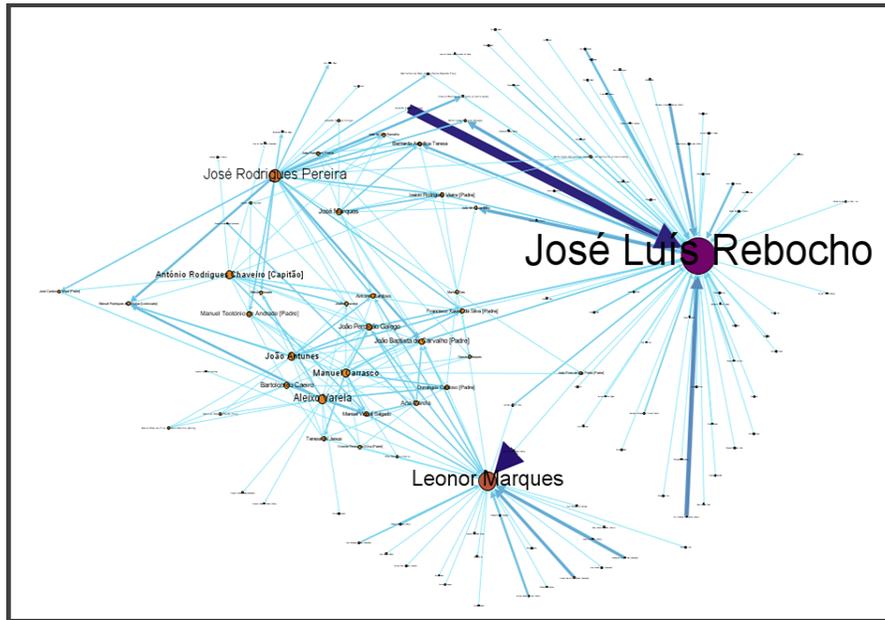


Figura 5. Testemunhas utilizadas nos concelhos de Arraiolos.

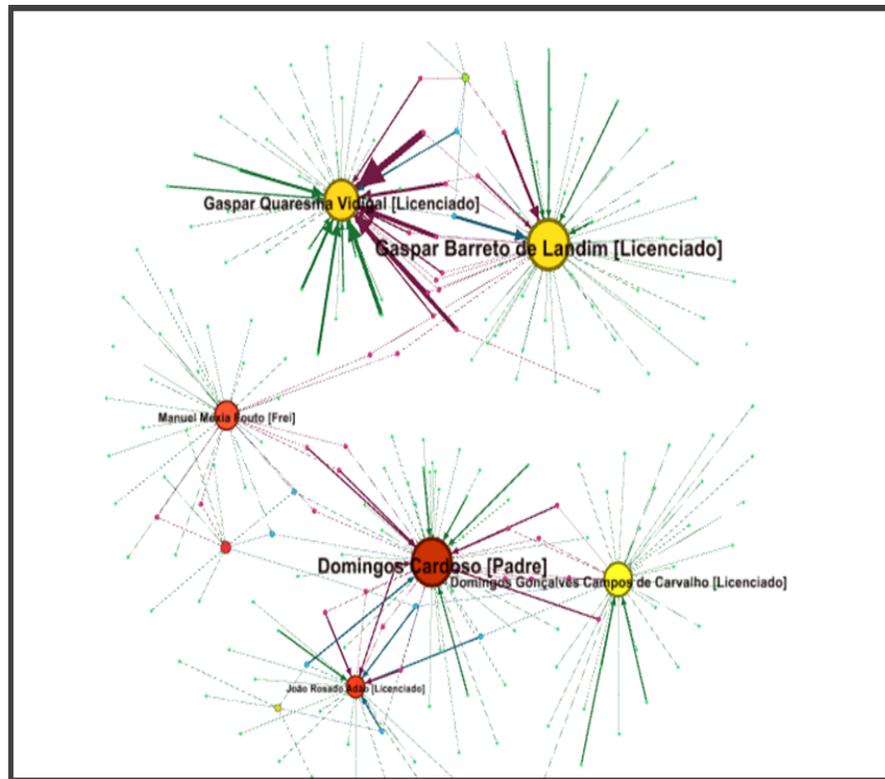


Figura 6. Comissários mais ativos no concelho de Arraiolos.

4 Conclusões e Trabalho Futuro

Pretendemos com o presente trabalho mostrar o interesse do tratamento e análise dos dados qualitativos, de modo a possibilitar a sua utilização na análise de redes sociais.

A preparação e manipulação dos dados efetuadas até ao momento permitiu-nos confirmar dois eventos relacionais que estatisticamente já se encontravam analisados. Este facto permite, pois, validar a atuação desta rede. No decorrer do nosso trabalho, e em estreita colaboração com os membros do nosso projeto, foi também possível identificar e corrigir algumas situações anómalas. Estas têm de ser feitas automaticamente, pois com grandes números quem introduz perde facilmente o controlo dos dados.

Relativamente à escolha do *software*, esta recaiu sobre o *GEPHI*, em detrimento do *PAJEK*. Com efeito, aquele *software* revelou-se mais adequado aos nossos objetivos, possibilitando uma integração mais amigável na aplicação que nos encontramos a desenvolver.

Em termos de futura investigação, consideramos pertinente trabalhar na produção de redes dinâmicas do ponto de vista da variação cronológica, dado que na nossa investigação nos deparamos constantemente com esta última, a par do caráter fragmentário dos dados que exploramos. Será também alvo do nosso trabalho a exploração do parentesco horizontal como rede, pois consideramo-lo muitas vezes mais importante que o parentesco vertical.

Uma das nossas metas é ainda construir uma aplicação que permita a adequação entre a base de dados prosopográfica *SPARES* e o *software* de redes *GEPHI*, tornando possível que qualquer utilizador de Ciências Sociais, e como tal menos familiarizado com a Estatística e a Informática, possa efetuar rápida e facilmente uma análise à rede social que estuda.

Referências

1. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks. In Proceedings of the Third International ICWSM Conference. California, USA. (2009) 361–362
2. Batagelj, V., Mrvar, A.: Pajek: Program for Analysis and Visualization of Large Networks. Reference Manual List of commands with short explanation version 2.00. University of Ljubljana. Slovenia. (2010)
3. Borgatti, S. P., Mehra, A., Brass, D. J., Labianca, G.: Network Analysis in the Social Sciences. *Science*. **323** (2009) 892–895
4. Lazer D. et al.: Life in the Network: the Coming Age of Computational Social Science. *Science*. **323** (2009) 721–723
5. Nooy, W, Mrvar, A, Batagelj, V.: Exploratory Network Analysis with Pajek. Cambridge University Press. New York. (2005)
6. Snijders, T.A.B., Steglich, C.E.G., van de Bunt, G.G.: Introduction to Actor-Based Models for Network Dynamics. *Social Networks*. **32** (2010) 44–60