

Das Bases de Dados Prosopográficas à Análise de Redes: Ensaio de Aplicação a Dados Históricos

Albertina Ferreira¹ Carlos Caldeira² Fernanda Olival³

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

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

³ Universidade de Évora; email: mfo@uevora.pt

Resumo As redes sociais são actualmente uma realidade incontornável e abrangente, fundamentais para melhor compreender a sociedade. Este trabalho pretende, em termos globais, iniciar uma análise das relações existentes entre os diversos protagonistas dos processos de Familiaturas do Santo Ofício. Em termos concretos, aborda-se a investigação e o conhecimento reunidos no âmbito das redes, das ontologias e de data mining. Por um lado, focalizamo-nos no interesse suscitado pela teoria dos grafos, com destaque para as medidas de centralidade, salientamos a importância da visualização das redes e das várias dinâmicas sociais. Por outro lado, referimos a importância do desenvolvimento de ontologias e da aplicação de data mining às redes sociais. No ensaio preliminar aplica-se uma das medidas de centralidade, o grau, ao nosso caso de estudo.

1 Introdução

Durante os últimos anos o estudo de teorias de redes nas ciências físicas e sociais tem sido uma área pela qual os investigadores demonstram grande interesse. Para os cientistas sociais, a teoria das redes tem possibilitado explicações para os fenómenos sociais, numa ampla variedade de disciplinas que vão desde a psicologia à economia [3].

Newman *et al.* [9] comentam que as redes estão em toda parte e que problemas dinâmicos estão na vanguarda da pesquisa em rede, onde há muitas perguntas ainda sem resposta. Mais recentemente Lazer *et al.* [8] referem que vivemos a vida em rede.

Embora os autores anteriormente focados considerem essencialmente redes a funcionar na actualidade, grande parte dos estudos que realizam poderão ser estendidos a outras épocas, bem como a outras sociedades. Neste trabalho propomo-nos estudar de que modo se relacionavam os diversos intervenientes nos processos de Familiaturas do Santo Ofício. Este estudo será enquadrado no âmbito mais lato das redes sociais e desenvolvido no âmbito do projecto aprovado e financiado pela FCT: PTDC/HIS-HIS/118227/2010 – Grupos intermédios em Portugal e no Império Português: as familiaturas do Santo Ofício (c. 1570-1773) – Instituição sede: CIDEHUS⁴.

⁴ Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora

2 Estado da Arte

Uma das ideias que mais prevalece nas ciências sociais é a noção de que as pessoas constituem redes que representam as relações sociais e suas interações. Deste modo, a teoria das redes sociais oferece uma resposta para uma questão que é colocada desde a época de Platão: como funciona a sociedade? A teoria de redes também fornece explicações para muitos fenómenos sociais, que vão desde a criatividade individual até à rentabilidade das empresas [3].

Para Snijders *et al.* [12], a evolução nas redes sociais é um domínio de investigação com alguma complexidade. Como é que uma rede social evolui? Podemos encontrar leis e derivar modelos que explicam a sua evolução? Como é que as comunidades surgem numa rede social? Estas questões são só um exemplo de todas aquelas que se podem colocar neste âmbito de investigação e às quais podemos acrescentar: como é que redes, onde a variação temporal é um ponto crucial, admitem cortes cronológicos e aceitam dados fragmentários?

São muitas as descrições que se encontram sobre redes sociais. Andery *et al.* [1] referem que a maneira mais usual de apresentar redes sociais é utilizando grafos. Estes são diagramas constituídos por vértices e arestas que representam os indivíduos da rede e as relações que estes mantêm entre si.

A estrutura das redes pode ser analisada utilizando medidas de centralidade: grau (número de relações que um nó estabelece com os nós que lhe são adjacentes), proximidade (distância de um nó relativamente aos restantes nós na rede), intermediação (intermediação que um nó faz com os nós que com ele não se conectam directamente) [7].

A aplicação de ontologias pode ser um caminho para melhor organizar-se a informação. Em 1993, Gruber [5] definiu ontologias como “*explicit specification of a conceptualization*”. Refere ainda que para sistemas baseados em conhecimento, o que existe é exactamente o que pode ser representado. Desde modo descreve a ontologia de um programa a partir da definição de um conjunto de termos representativos. Embora não exista uma metodologia modelo para aplicar no desenvolvimento de uma ontologia este autor salienta que esta deve ser clara, objectiva, coerente e extensível.

No que respeita a metodologias aplicáveis à construção de uma ontologia Rautenberg *et al.* [11] referem cinco etapas: especificação (identificar o objectivo e o âmbito da ontologia); conceptualização (descrever um modelo conceptual); formalização (transformar o modelo conceptual num modelo formal); implementação (desenvolver a ontologia formalizada numa linguagem de representação apropriada); manutenção (actualizar e corrigir a ontologia sempre que seja necessário).

A organização da informação conseguida através de uma ontologia poderá ser valorizada, gerando conhecimento que posteriormente pode ser partilhado, quando se utilizam técnicas de *data mining*. Através destas é possível identificar padrões e tendências em repositórios com elevado número de registos. Fayyad *et al.* [4] definem assim *data mining* como: “... a step in the KDD process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of

patterns (or models) over the data". Estes autores referem ainda o interesse, para a procura de padrões, da aplicação de regras de classificação, regressões e análise de *clusters*. Para além destas, Han *et al.* [6] consideram também a análise de *outliers* como uma das funcionalidades de *data mining*.

Para Han *et al.* [6] a análise de *clusters* estuda e agrupa os dados que possuem características similares, permitindo a organização de observações numa hierarquia de classes que agrupam eventos semelhantes. A análise de *outliers*, referida por estes autores, investiga os dados que não apresentam um comportamento idêntico ao da maioria, revelando-se muitas vezes os eventos raros mais interessantes do que os eventos que regularmente ocorrem.

3 Procedimentos Metodológicos

O estudo em questão possui uma perspectiva longitudinal, tendo em vista que os dados a utilizar se encontram distribuídos por três séculos, recaindo o nosso estudo sobre um número bastante elevado de registos, os quais se encontram disponíveis na base de dados prosopográfica *SPARES*⁵, que actualmente tem 83.163 registos. Esta base de dados foi desenvolvida por Carlos Pampulim Caldeira no âmbito do projecto FCOMP-01-0124-FEDER-007360 – Inquirir da Honra: Comissários do Santo Ofício e das Ordens Militares em Portugal (1570 – 1773).

A ferramenta escolhida para construir e analisar a rede foi o *Pajek*. Na base desta escolha esteve, por um lado, a possibilidade desta aplicação poder explorar e manipular redes de grande dimensão [10] e, por outro, correr sobre o sistema Windows. Ele está disponível gratuitamente, para uso não comercial, e pode ser acedido a partir de: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> [2]. É utilizando este software que iremos construir a rede e analisá-la no que diz respeito a uma das medidas de centralidade referida no capítulo 2: grau.

Por possuir uma arquitectura modular e estar em constante desenvolvimento, a ferramenta *Protégé*, disponível em <http://protege.stanford.edu> foi o editor que escolhemos para futuramente utilizarmos na construções e desenvolvimento de ontologias. Esta ferramenta permite a implementação de diversas metodologias que possibilitam não só a definição de classes e de hierarquias como também a implementação de restrições ao nível das propriedades.

Consideramos que as tarefas de *data mining* serão fundamentais para o reconhecimento de padrões, concretamente no que respeita à análise de *clusters* e de *outliers*. Pensamos complementar a abordagem que será feita através do *Pajek* com a utilização da ferramenta *Weka*. Esta encontra-se disponível gratuitamente em <http://www.cs.waikato.ac.nz/ml/weka/>.

4 Ensaio Preliminar

Iniciou-se este ensaio com uma pesquisa à base de dados *SPARES*. Nesta trataram-se os dados gerando meta-dados, de modo que estes pudessem ser analisados no

⁵ Sistema Prosopográfico de Análise de Relações e Eventos Sociais

Pajek. Utilizando esta ferramenta criou-se a rede que pode ser observada na Figura 1. A observação global desta rede não é impossível, mas, como se pode

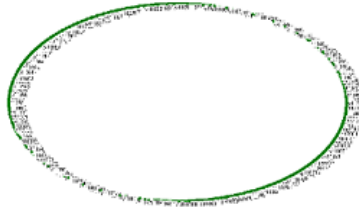


Figura 1. Rede visualizada com a visualização Layout Circular/Random.

constatar trata-se de uma visualização a uma escala pouco esclarecedora. Continuámos este ensaio considerando duas hipóteses. Na primeira considerámos o tipo de relação que existia entre os diversos protagonistas. Na segunda, fizemos uma restrição no que diz respeito à data inicial (este é um dado sempre conhecido, nem que seja por aproximação). Concluimos ambas as hipóteses fazendo um estudo sobre uma das medidas de centralidade anteriormente referidas: grau.

4.1 Exploração de dados por tipo de relação

Na base de dados *SPARES* encontram-se referidos diversos tipos de relações. Apresentamos neste ensaio os procedimentos realizados para uma relação de “Patrocínio” (quando um indivíduo intervém de forma explícita a favor de outro). Apresenta-se na Figura 2 o resultado da visualização desta relação no *Pajek*. Pode-se observar que, para este tipo de relação, cinco (72%) dos indivíduos têm

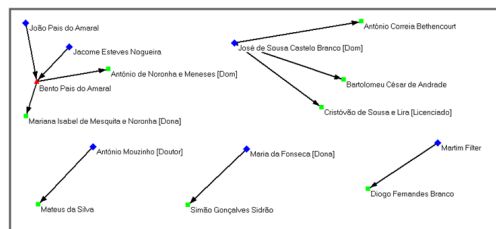


Figura 2. Rede criada no *Pajek* para uma relação de Patrocínio.

apenas um patrocínio, um (14%) indivíduo patrocina dois indivíduos e um (14%) indivíduo patrocina três indivíduos. De um total de 15 indivíduos, apenas um (Bento Pais do Amaral) patrocina e é simultaneamente patrocinado (7%).

Considerando que o grau é o número de relações de cada um dos comissários relativamente às testemunhas, detectaram-se três situações possíveis: 5 comissários possuem grau 1; 1 comissário possui grau 2; 1 comissário possui grau 3. Podemos assim constatar que a testemunha (que também é comissário) Bento Pais do Amaral é *prominent*, possui grande prestígio, e que o comissário Dom José de Sousa Castelo Branco é *influential* e possui uma forte ascendência sobre todos os outros [7].

4.2 Exploração dos dados limitando a data inicial

Para a segunda hipótese, considerámos os processos relativos ao século XVI, no que diz respeito à relação entre comissários e testemunhas. Pode-se observar na Figura 3 o resultado visualizado no *Pajek*. Da observação destes dados podemos

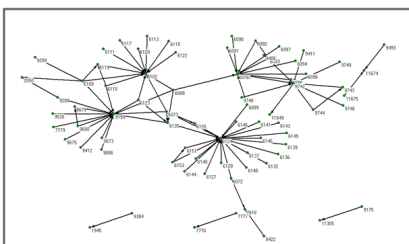


Figura 3. Rede criada no *Pajek* para uma data inicial pertencente ao século XVI.

retirar algumas ilações: existem 5 comissários com um número elevado de testemunhas; alguns comissários têm apenas uma testemunha; na maior parte das situações, as testemunhas encontram-se apenas relacionadas com um comissário; algumas testemunhas são simultaneamente comissários. Apresenta-se na Tabela 1 os resultados obtidos relativamente à medida de centralidade grau. Verifica-se,

Tabela 1. Número de testemunhas por graus

Grau	Número de testemunhas
1	51
2	8
3	2
4	1
5	2

assim, que 67% das testemunhas encontra-se relacionada com apenas um comissário, 11% com 2 comissários, 3% com 3 comissários, 1% com 4 comissários

e 3% com 5 comissários. Estas testemunhas totalizam 84% da rede, sendo nalgumas situações também comissários. Os restantes 16% da rede correspondem a comissários que não se encontram relacionados com nenhuma testemunha.

5 Conclusões e Trabalho Futuro

Este estudo pretendeu iniciar-nos na análise de redes que permita estudar as relações que existiam entre os diversos protagonistas nos processos de recrutamento de agentes para o Santo Ofício.

Como trabalho futuro, pretendemos por um lado, produzir redes dinâmicas em função de variáveis cronológicas, uma vez que a variação temporal, bem como o carácter fragmentário dos dados são uma constante nos dados que exploramos. Por outro lado desejamos explorar o parentesco horizontal como rede, pois este conhecimento é muitas vezes mais importante do que conhecer o parentesco vertical que actualmente se encontra implementado. É também nossa intenção empregar *data mining* e ontologias para a determinação de padrões que possam ser úteis e interpretáveis. Por último, é nosso desejo encontrar alternativas de interoperabilidade entre as várias ferramentas anteriormente referidas.

Referências

1. Andery, G. F., Lopes, A. A., Minghim, R.: Exploração visual multidimensional de redes sociais. 2nd International Workshop on Web and Text Intelligence. São Carlos. (2009) 1–9
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. Fayyad, U., Piatetsky-Shapiro, G.; Smyth, P.: From data mining to knowledge discovery: an overview. *AI Magazine*. **17(3)**: (1996) 37–54.
5. Gruber, T.: A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*. **5(2)**: (1993) 199–220.
6. Han, J.; Kamber, M.; Pei, J.: *Data Mining-Concepts and Techniques*. Third Edition. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers. (2011)
7. Hanneman, R. A., Riddle, M.: *Introduction to social network methods*. University of California, Riverside. (2005) <http://faculty.ucr.edu/~hanneman/>
8. Lazer D. et al.: Life in the Network: the Coming Age of Computational Social Science. *Science*. **323** (2009) 721–723
9. Newman, M. E. J., Barabási, A., Watts, D. J.: *The Structure and Dynamics of Networks*. (2006) <http://press.princeton.edu/chapters/s8114.html>
10. Nooy, W, Mrvar, A, Batagelj, V.: *Exploratory Network Analysis with Pajek*. Cambridge University Press. New York. (2005)
11. Rautenberg, S.: Uma Metodologia para o Desenvolvimento de Ontologias. *Revista Ciências Exatas e Naturais*. **10(2)**: (2008) 237–262
12. 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