

Dados qualitativos, prosopografia e análise de redes

Albertina Ferreira¹, Carlos Caldeira², Fernanda Olival³

¹Escola Superior Agrária, Instituto Politécnico de Santarém, albertina.ferreira@esa.ipsantarém.pt

²Departamento de Informática, Universidade de Évora, ccaldeira@di.uevora.pt

³Departamento de História, Universidade de Évora, mfo@uevora.pt

1 Introdução

Para Snijders *et al.* (2010:149) 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?

Guarino *et al.* (2009:2) referem que a criação de ontologias permite modelar formalmente a estrutura de um sistema, permitindo alcançar novos conhecimentos. É neste sentido que pretendemos propor a construção de uma ontologia ao nosso caso de estudo.

Para esta descoberta e gestão de conhecimento contribuirá também a aplicação de *data mining*. A identificação de padrões e tendências em bases de dados com um número elevado de registos reforçam o interesse da aplicação de *data mining* às redes sociais. Günnemann *et al.* (2011: 1363) referem essa identificação de padrões, destacando a importância da interpretação dos resultados obtidos.

Para concretizar os diversos componentes do nosso estudo, nomeadamente ontologias, *data mining* e redes sociais, 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 projeto aprovado e financiado pela FCT¹: PTDC/HIS-HIS/118227/2010 – Grupos intermédios em Portugal e no Império Português: as familiaturas do Santo Ofício (c. 1570-1773) – Instituição sede: CIDEHUS ².

2 Objetivos

O presente estudo reveste-se de importância na investigação histórica, pois, quando conseguimos perceber o passado podemos contribuir para uma adequada explicação e construção de modelo da sociedade de determinada época.

É nosso objetivo que este trabalho possa ser um elemento importante no estudo dos mais diversos fenómenos sociais em que a variável tempo tenha um peso considerável. Pretendemos fazer um ensaio à sociedade portuguesa dos séculos XVI, XVII e XVIII, nomeadamente no que diz respeito aos processos de familiaturas do Santo Ofício. Ter-se-á em consideração que ao estudarmos sociedades do passado estamos a tratar de dados fragmentários e complexos. Assim, as bases de dados prosopográficas devem admitir a possibilidade de correção temporal.

Os objetivos principais do nosso projeto são:

- Estudar as relações que existiam entre os diversos protagonistas, visualizá-las graficamente e perceber a importância de alguns elos e intervenção para que os atores sociais atinjam metas que lhes são vantajosas;
- Construir uma ontologia que, por um lado, permita acrescentar conhecimento ao nosso caso de estudo e por outro possa ser adequada a outras situações. Para concretizar este objetivo recorrer-se-á à aplicação de *data mining*.

Como objetivos secundários, poderemos apontar os seguintes:

- Realizar a adequação entre a base de dados prosopográfica e o *software* de redes. Esta base de dados permite armazenar informação sobre eventos biográficos e relacionais, analisar genealogias e cartografar dados. Falta-lhe permitir o estudo das redes de relações entre os atores sociais;
- Tornar o *software* de redes amigável para qualquer utilizador de Ciências Sociais e como tal menos familiarizado com a Estatística e a Informática;

¹ Fundação para a Ciência e a Tecnologia

² Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora.

- Produzir redes dinâmicas em função de variáveis cronológicas;
- Explorar o parentesco horizontal como rede.

Espera-se, com este trabalho, não só contribuir para uma melhor compreensão de como intervinham as redes nos grupos intermédios, mas também deixar uma porta aberta a futuros ensaios noutras áreas distintas, em que a variável tempo seja um fator a considerar.

3 Metodologia

O estudo em questão possui uma perspetiva 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 *SPARES*³, 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 *SPARES* de modo que estes possam ser manipulados por *softwares* de redes. Escolhemos o *PAJEK* para construirmos e analisarmos a nossa rede porque 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/>. Embora o *PAJEK* tenha sido a aplicação elegível pensamos ainda vir a utilizar neste contexto outros *softwares* que se adequem ao caso de estudo.

A construção de uma ontologia permitirá organizar melhor todos os dados disponibilizados e conseqüentemente incrementar o conhecimento já existente acerca dos processos de Familiaturas do Santo Ofício. Por possuir uma arquitectura modular e estar em constante desenvolvimento, a ferramenta *PROTÉGÉ* foi o editor que escolhemos para utilizar na construção 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. Esta ferramenta, bem como tutoriais, documentação ou livrarias, encontra-se disponível em <http://protege.stanford.edu>.

Na utilização de *data mining*, nomeadamente no reconhecimento de padrões, pensamos utilizar a ferramenta *WEKA*. Em <http://www.cs.waikato.ac.nz/ml/weka/> pode-se aceder a toda a informação relativamente a esta ferramenta. A escolha recaiu nesta ferramenta por ser *open source* e gratuita.

4 Resultados

Como já referimos na metodologia efetuou-se a automatização da extração dos dados da base de dados *SPARES*. Para atingir este objetivo foi necessário realizar diversos tipos de pesquisas que permitissem por um lado criar novas tabelas ou acrescentar registos e/ou atributos a tabelas já existentes. Associado a estas pesquisas criaram-se também alguns procedimentos em VBA. Ilustra-se na **Figura 1** um exemplo dos dados antes da sua manipulação e na **Figura 2** os dados já preparados para serem lidos no *PAJEK* (ou noutro *software* que possa utilizar este formato).

Nome de P1	Código de P1	Relação	Código de P2	Nome de P2
Luis de Araujo 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
Luis de Araujo de Barros	9199	Consulta favorável no CG	6123	António de Matos de Noronha [Dom]
Luis de Araujo 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

```

#Vertices 6
1 "António de Matos de Noronha [Dom]" ic Blue bc Blue
2 "Diogo de Sousa [Doutor]" ic Blue bc Blue
3 "Luis de Araujo de Barros" box ic Red bc Red
4 "Marcos Teixeira" ic Blue bc Blue
5 "Martim Afonso de Melo" box ic Red bc Red
6 "Salvador de Mesquita" box ic Red bc Red

#Arcs
4 3 1
2 3 1
1 3 1
4 5 1
4 6 1
2 6 1
1 6 1

```

Figura 2

Na **Figuras 3** e na **Figura 4** pode-se observar o interface que permite a extração dos dados presentes na base de dados *SPARES*, bem como a utilização destes pelo *PAJEK*.

³ Sistema Prosopográfico de Análise de Relações e Eventos Sociais.



Figura 3



Figura 4

A rede criada, no *PAJEK*, a partir das escolhas feitas no interface da **Figura 4** pode ser consultada na **Figura 5**.

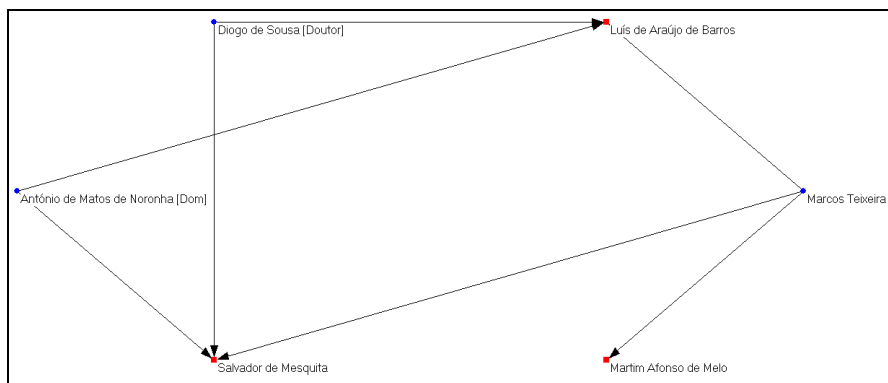


Figura 5

5 Conclusões

Os dados analisados encontram-se dependentes temporalmente dos processos de Familiaturas do Santo Ofício. Neste sentido, ao estudar a progressão temporal dos acontecimentos é possível estudar o impacto desses acontecimentos sobre a sociedade Portuguesa nos séculos XVI, XVII e XVIII.

Neste momento a aplicação ainda se encontra em aperfeiçoamento, nomeadamente no que respeita à interligação com o *PAJEK*.

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, temos como objetivo explorar o parentesco horizontal como rede, pois este conhecimento é muitas vezes mais importante do que conhecer o parentesco vertical que atualmente se encontra implementado.

É também nossa intenção empregar *data mining* e ontologias para a determinação de padrões que possam ser úteis e interpretáveis, facilitando o uso das ferramentas de rede por utilizadores pouco versados em Informática.

Por último, é nosso desejo encontrar alternativas de interoperabilidade entre as várias ferramentas anteriormente referidas.

6 Referências

- Guarino, N.; Oberle, D.; Staab, S. 2009. What is an Ontology? S. Staab; R. Studer (eds.), Handbook on Ontologies, Second Edition. *International handbooks on information systems*. Berlin: Springer Verlag. 1-17.
- Günemann, S.; Färber, I.; Müller, E.; Assent, I.; Seidl, T. 2011. External Evaluation Measures for Subspace Clustering. *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM 2011)*. Glasgow, UK. 1363-1372.
- Snijders, T.A.B.; Steglich, C.E.G.; van de Bunt, G.G. 2010. Introduction to Actor-Based Models for Network Dynamics. *Social Networks* 32: 44-60.

Palavras-chave: *data mining*, ontologias, redes sociais, sequências temporais.