

Redes de Recomendações na Seleção de Recursos Humanos

David Fernandes

Mestre em Tecnologias e Sistemas Informáticos WEB, Universidade Aberta
902006@estudante.uab.pt, david.paiva.fernandes@gmail.com

Luís Cavique

Universidade Aberta, DCeT
luis.cavique@uab.pt

Resumo

Um dos principais desafios encontrados num processo de recrutamento de recursos humanos prende-se com a dificuldade prática de analisar de forma objetiva todos os candidatos. Neste trabalho mostra-se a aplicabilidade das redes de recomendações profissionais entre pares no âmbito de processos de recrutamento de recursos humanos. Para o efeito utilizamos a rede social LinkedIn e em particular as recomendações efetuadas por um conjunto de profissionais. Geramos uma rede de recomendações por especialidade, em que os vértices são os profissionais analisados e os arcos são as relações de recomendação. Sobre a rede de recomendações aplica-se o algoritmo PageRank para ordenar cada profissional por especialidade. A análise para várias especialidades é realizada através da avaliação multicritério, onde é aplicado o método TOPSIS.

Palavras-chave: recursos humanos, rede de recomendações, PageRank, TOPSIS.

Title: Recommendation networks in human resource selection

Abstract

One of the main challenges in a human resources recruitment process is the difficulty of analyzing objectively all candidates. This paper shows the application of networks of professional recommendations among peers in human resources recruitment processes. For this purpose we use the LinkedIn social network and in particular the recommendations made by a subset of professionals. We generate a network of recommendations by specialty, where the vertices are the professionals and the arcs are the recommendation relationships. Based on the network of recommendations we apply the algorithm PageRank to order of each professional by specialty. The analysis for several specialties is performed using multi-criteria evaluation, where the TOPSIS method is applied.

Keywords: human resources, recommendations network, PageRank, TOPSIS.

1 – Introdução

Um dos principais problemas na seleção de recursos humanos é o elevado número de CV que são recebidos no processo de recrutamento. Os CV carecem de verificação em tempo útil das competências, das referências e do percurso profissional.

Gusdorf et al. (2008) refere estudos que indicam que, nos Estados Unidos, 40% dos candidatos mentem sobre o seu percurso profissional e académico e estima-se que 30% dos candidatos colocam algum tipo de falsidade nos seus CV. Um inquérito a alunos universitários refere que 95% dos inquiridos afirmam que mentiriam para conseguir trabalho e 41% reconhecem já o ter feito em algum momento. Num outro inquérito, este efetuado a gestores de topo, conclui-se que 15% dos entrevistados admitem ter falsificado o seu CV.

Neste contexto, o conceito de recomendação aliado ao da reputação (Gkorou et al., 2015) podem ter um papel importante na classificação de elementos numa rede de relações e, no caso particular do recrutamento, podem constituir um indicador muito útil na avaliação de um potencial colaborador.

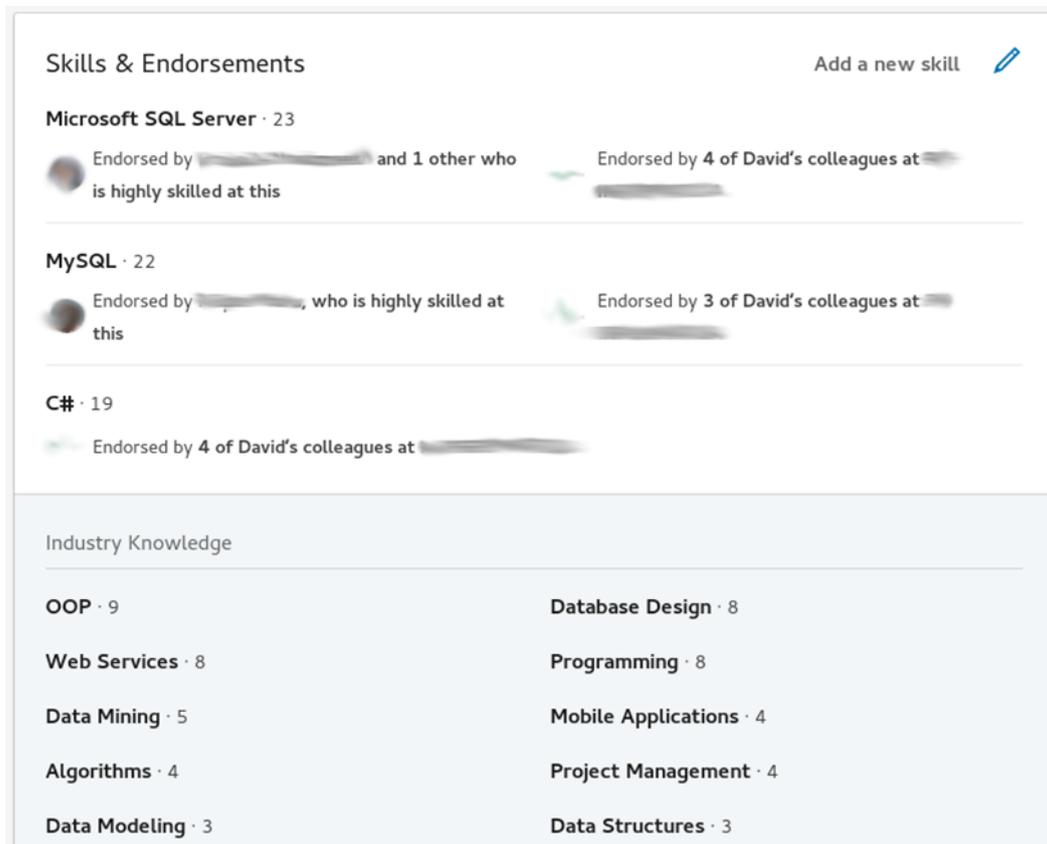


Figura 1 – Aspeto do quadro de competências e recomendações do LinkedIn

Para criar uma rede de recomendações, neste trabalho iremos utilizar os dados disponíveis na rede social LinkedIn. Na Figura 1, o candidato David é recomendado por 23 membros na especialidade de SQL-Server e por 22 membros na especialidade de MySQL

Com base na rede de recomendações, o objetivo deste trabalho é definido da seguinte forma: existindo a necessidades de recrutar p profissionais com competências em s especialidades, de um conjunto de n candidatos, sendo $p \leq n$, pretende-se encontrar os p candidatos com o melhor perfil para o desempenho das s funções.

Este trabalho é uma extensão do trabalho (Fernandes 2015) onde é utilizado o algoritmo PageRank para ordenar candidatos no recrutamento de recursos humanos. Este trabalho reutiliza ainda o trabalho desenvolvido por (Fernandes 2018) onde se utiliza o algoritmo TOPSIS para ordenar candidatos com duas ou mais especialidades.

A contribuição mais relevante deste trabalho prende-se com a apresentação de um método inovador que utiliza de redes de recomendações para seriar candidatos com um conjunto de competências, utilizando dois algoritmos bem conhecidos, o PageRank e o TOPSIS.

A estrutura do artigo é a seguinte. Na secção 2 é referido o trabalho relacionado, dando realce aos métodos e algoritmos utilizados em processos de recrutamento. Na secção 3 é apresentado o modelo proposto, referindo (i) a extração de dados, (ii) a ordenação dos candidatos com utilização do algoritmo PageRank e a (iii) utilização de técnicas multicritério, utilizando o algoritmo TOPSIS, quando são requeridas duas ou mais especialidades. Na secção 4 é apresentado um caso de uso com dados reais. Na secção 5 são apresentadas as conclusões.

2 – Trabalho relacionado

A web veio alterar radicalmente os processos de gestão de recursos humanos tradicionais, principalmente como meio de comunicação (Kessler et al. 2009). De facto, através do potencial de comunicação da web, empregadores e candidatos ficaram mais próximos e consequentemente viram agilizados todos os processos de recrutamento, seleção e colocação.

Capiluppi et al. (2013) referem um estudo de Barnes e Mattson (2009) que conclui que 48% das companhias da lista Inc. 500 utilizaram *sites* e redes sociais para recrutamento e avaliação de candidatos e um inquérito, efetuado a 115 PME, em que se verificou que 78% das empresas inquiridas utilizaram também redes sociais nos seus esforços de recrutamento. Os mesmos autores indicam ainda as três redes sociais mais utilizadas: LinkedIn, Facebook e Jobster.

Inúmeros sistemas surgiram para apoio ao mercado do recrutamento, como sejam: seguimento e avaliação de candidatos, sistemas de gestão de relacionamento de candidatos, sistemas de referência social e ferramentas de entrevista, mas o tipo de sistema mais interessante neste mercado são as ferramentas de deteção de recursos (Bersin 2012).

Strohmeier e Piazza (2013), de uma forma mais abrangente, analisam as potencialidades do data-mining neste tipo de problema e referem como áreas de utilização: a seleção de candidatos, a previsão do retorno do colaborador, o recrutamento e detecção de competências, planeamento de carreiras, desenvolvimento e planeamento de custos de RH, previsão e avaliação do desempenho do colaborador.

Bakar e Ting (2011) aplicam redes Bayesianas a um problema com que se deparam os recrutadores na elaboração dos requisitos para uma contratação: a escolha das competências transversais não técnicas, as chamadas soft skills, mais adequadas a uma determinada necessidade de recrutamento. Domeniconi et al. (2016) usam análise semântica latente, uma técnica de processamento de linguagem natural, para encontrar semelhança semântica entre as competências de membros do LinkedIn.

Esta investigação tem a sua origem no desejo de aprofundar um trabalho exploratório realizado no âmbito da componente curricular deste mestrado (Fernandes 2015) e apresentado na InforAberta 2015, Jornadas de Informática da Universidade Aberta. Neste trabalho introduz-se a utilização do algoritmo PageRank a uma rede de recomendações extraída da informação sobre Skills/Endorsements do LinkedIn. Neste trabalho propõe-se ainda o desenvolvimento de um método que permita a dedução de informação adicional através da análise de correlações entre especialidades.

O algoritmo PageRank aplicado a redes de recomendações é também utilizado por Pérez-Rosés et al. (2016) e Pérez-Rosés e Sebé (2017). Nestes artigos descrevem a dedução de um indicador de autoridade de recomendação baseado na aplicação de um algoritmo de ordenação como o PageRank. Previamente à utilização do PageRank, os autores enriquecem informação contida no grafo de uma dada especialidade com novos arcos de peso ponderado através do cálculo de correlações entre diferentes especialidades.

Na Secção 3 apresentamos o modelo proposto (Fernandes 2018) que numa primeira fase ordena os candidatos por especialidade e de seguida usa técnicas multicritério para combinar duas ou mais especialidades.

3 - Modelo proposto PageRank-TOPSIS

O modelo proposto, intitulado PageRank-TOPSIS, ordena os vários candidatos com diferentes especialidades, desenvolvendo-se em três fases:

- i. Extração de dados da rede social LinkedIn;
- ii. Ordenação dos candidatos por especialidade utilizando o algoritmo PageRank;
- iii. Combinação de múltiplas especialidades utilizando o algoritmo TOPSIS.

3.1 – Extração dos dados do LinkedIn

A informação que nos interessa neste trabalho, e que infelizmente não é disponibilizada pela API do LinkedIn, é a lista de aprovações ou de competências reconhecidas. Estas aprovações não são mais do que referências de um membro às capacidades de um outro sobre um determinado assunto.

Por exemplo: o João indica que a Patrícia possui conhecimentos de Java, e indica que o Manuel conhece Redes Neurais. Na rede de recomendações é composta por nós e relações. Para cada especialidade existe uma rede de recomendações. Os nós, ou atores, são os candidatos João, Patrícia e Manuel. Na rede da especialidade Java temos a relação (João, Patrícia) e na rede de recomendação da especialidade Redes Neurais temos a relação (João, Manuel).

Note-se que estas recomendações não possuem qualidade. Não se trata portanto de avaliações mas apenas referências positivas, já que não se podem fazer referências negativas. A rede indica que determinado membro refere outro membro como conhecedor numa determinada especialidade.

O LinkedIn disponibiliza uma API de acesso controlado que permite aceder a alguma da informação do membro. No entanto, a API não disponibiliza informação relativa à rede de recomendações, pelo que esta teve de ser extraída das páginas dos membros diretamente a partir do HTML.

Na medida em que o processo de extração de dados, não totalmente automático, iria acarretar algum trabalho manual repetitivo, optou-se por criar um serviço WEB REST a ser invocado pelo script JS e que trataria de gravar a informação recebida numa pequena base de dados mySQL criada para o efeito.

A arquitetura deste sistema de extração de dados dos LinkedIn está representado na Figura 2.

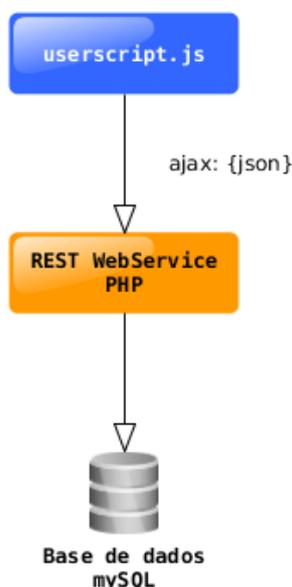


Figura 2 - Arquitetura para extração de dados do LinkedIn

Com o objetivo de acelerar o processo de transferência da informação coletada pelo userscript.js para um local de onde pudesse ser posteriormente consumida, foi criado um serviço WEB REST em PHP sobre uma base de dados mySQL.

Para cada membro, a informação coletada é formatada numa estrutura JSON e enviada para o serviço WEB REST de armazenamento. Na Figura 3 apresenta os dados das recomendações do membro “dfernandes” na especialidade de “Microsoft SQL Server”.

```
{
  "id": "/in/david/",
  "area": "Microsoft SQL Server",
  "endorsers": [
    { "id": "/in/jorge/",
      "name": "Jorge Amado"
    },
    { "id": "/in/falcao/",
      "name": "João Falcão"
    },
    ...,
  ]
}
```

Figura 3 – Dados das recomendações de um membro numa determinada especialidade

3.2- Algoritmo PageRank

O algoritmo conhecido como PageRank (Brin e Page, 1998) responde a um problema de *Link-based Object Ranking* especificamente aplicado a páginas *Web*. A ideia passa por calcular um indicador numérico (o valor de PageRank da página) indicativo da relevância de uma página WWW explorando apenas a estrutura de ligações da *Web* e em especial a quantidade e qualidade das ligações que lhe são feitas.

Sendo a WWW uma rede, facilmente se percebe que a sua estrutura corresponde a um grafo no qual os vértices são as páginas e as arestas as hiperligações entre elas.

Da mesma forma, ressalta aqui a semelhança em termos de estrutura entre a rede de páginas WWW e uma rede de recomendações, razão pela qual nos parece um algoritmo aplicável, também, no nosso problema.

O algoritmo PageRank pode ser expresso pela seguinte expressão recursiva, em que N é o número de páginas P_1, P_2, \dots, P_n , $M(P_i)$ o conjunto das páginas que referenciam P_i , $L(P_j)$ o número de referências da página P_j e d o fator de amortecimento que varia entre 0 e 1.

$$PR(p_i) = \frac{1 - d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

A versão iterativa do algoritmo PageRank considera o grau de entrada de cada nó e o valor de PageRank dos nós cujas arestas o referenciam.

A package *igraph* (Csardi e Nepusz, 2006) inclui várias funções para análise e visualização de redes e possui implementações em R, Python e C/C++. Neste trabalho é utilizada a implementação em R do algoritmo PageRank disponibilizada na função `page_rank`:

```

1: page_rank(
2:   graph,
3:   algo = c("prpack", "arpack", "power"),
4:   vids = V(graph),
5:   directed = TRUE,
6:   damping = 0.85,
7:   personalized = NULL,
8:   weights = NULL,
9:   options = NULL)

```

3.3 – Avaliação multicritério com TOPSIS

A avaliação multicritério tem como base uma matriz de decisão. A matriz bidimensional Critérios versus Alternativas tem m colunas com critérios e n linhas com alternativas, conforme representado na Figura 4. O resultado da avaliação multicritério é a ordenação das alternativas.

Alternativas	Critérios					
	C_1	C_2	...	C_j	...	C_m
A_1	W_{11}	W_{12}	...	W_{1j}	...	W_{1m}
A_2	W_{21}	W_{22}	...	W_{2j}	...	W_{2m}
...
A_i	W_{i1}	W_{i2}	...	W_{ij}	...	W_{im}
...
A_n	W_{n1}	W_{n2}	...	W_{nj}	...	W_{nm}

Figura 4 – Matriz de decisão multicritério

No nosso caso particular os critérios vão corresponder às especialidades escolhidas pela equipa de recrutamento e as alternativas correspondem aos candidatos. O resultado pretendido da avaliação multicritério é a ordenação dos candidatos.

Existem vários algoritmos na teoria da decisão multicritério que se podem dividir em duas classes: os compensatórios e os não-compensatórios. Os não-compensatórios defendem que um mau desempenho de um determinado critério não pode ser compensado por bons desempenhos em outros critérios.

Neste trabalho recorreremos a algoritmos compensatórios que permitem a agregação dos critérios e a ordenação das alternativas. A escolha recaiu sobre o algoritmo TOPSIS.

O algoritmo TOPSIS (Hwang, Yoon 1981) (*Technique for order preference by similarity to ideal solution*) é uma técnica de avaliação multicritério que se baseia no princípio de que, de um conjunto de alternativas de vários critérios, aquela que se constituirá como preferencial será a que se encontre à menor distância euclidiana da solução ideal positiva e à maior distância euclidiana da solução ideal negativa, ou anti-ideal. Assume-se que os critérios podem tomar valores de utilidade linearmente crescentes ou decrescentes.

Na Figura 5 apresenta-se um exemplo para dois critérios, X1 e X2, que se pretende que sejam maximizados. A solução ideal A+ inclui os melhores valores em diferentes critérios de todas as alternativas e a solução anti-ideal A- é composto pelos piores valores de todas as alternativas. Para cada alternativa A_i vamos procurar as distâncias euclidianas à solução ideal, d_i +, e à solução anti-ideal, d_i -. Na figura as distâncias d_i + e d_i - são apresentadas para as alternativas A1 e A2.

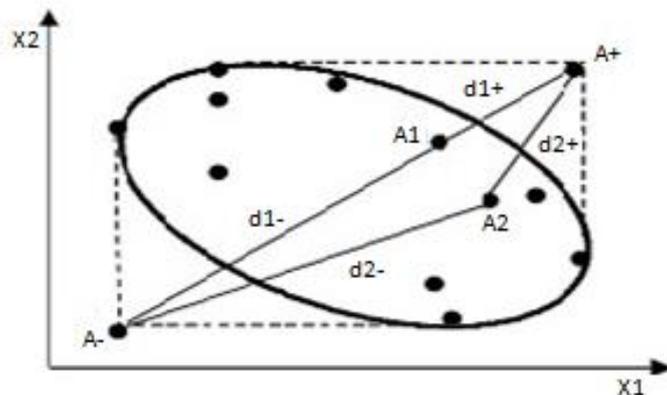


Figura 5 - Solução ideal e anti-ideal para um exemplo com dois critérios

O cálculo da proximidade relativa de cada alternativa, R_i , à solução ideal é dada pela expressão:

$$R_i = \frac{d_{i-}}{d_{i+} + d_{i-}}, 0 < R_i < 1, \quad i = 1, 2, \dots, m$$

A ordenação por ordem decrescente de R_i devolve a ordem de preferência das alternativas.

O package R inclui a implementação do algoritmo TOPSIS. De notar que a algoritmo prevê que nem todos os atributos possuam a mesma importância. Para isso, permite a referência de um vetor de pesos, pesos esses que indicam numericamente a importância de cada um dos atributos a considerar. A função TOPSIS aceita 7 argumentos:

```

1: TOPSIS (
2:   performanceTable,
3:   criteriaWeights,
4:   criteriaMinMax,
5:   positiveIdealSolutions = NULL,
6:   negativeIdealSolutions = NULL,
7:   alternativesIDs = NULL,
8:   criteriaIDs = NULL)

```

4 – Caso de uso

Nesta secção apresenta-se um caso de uso que exemplifica a aplicação do método PageRank-TOPSIS. É apresentada a visualização de uma rede de recomendação, calculado o PageRank para a rede e exemplificada a ordenação de candidatos para um conjunto de especialidades.

Os dados utilizados neste trabalho foram obtidos por extração automática de dados do LinkedIn e correspondem a 6579 recomendações efetuadas a 15 membros num total de 177 membros em 130 especialidades.

4.1 – Ordenação dos candidatos por especialidade

Com o objetivo de mostrar graficamente algumas das características do algoritmo PageRank, nesta secção apresentamos resultados de cálculos realizados apenas sobre uma rede de recomendação por especialidade. De facto, a representação gráfica das redes de múltiplos critérios não seria facilmente perceptível. A Figura 6 mostra uma rede de recomendação dos seis candidatos com melhor PageRank, para especialidade SQL.

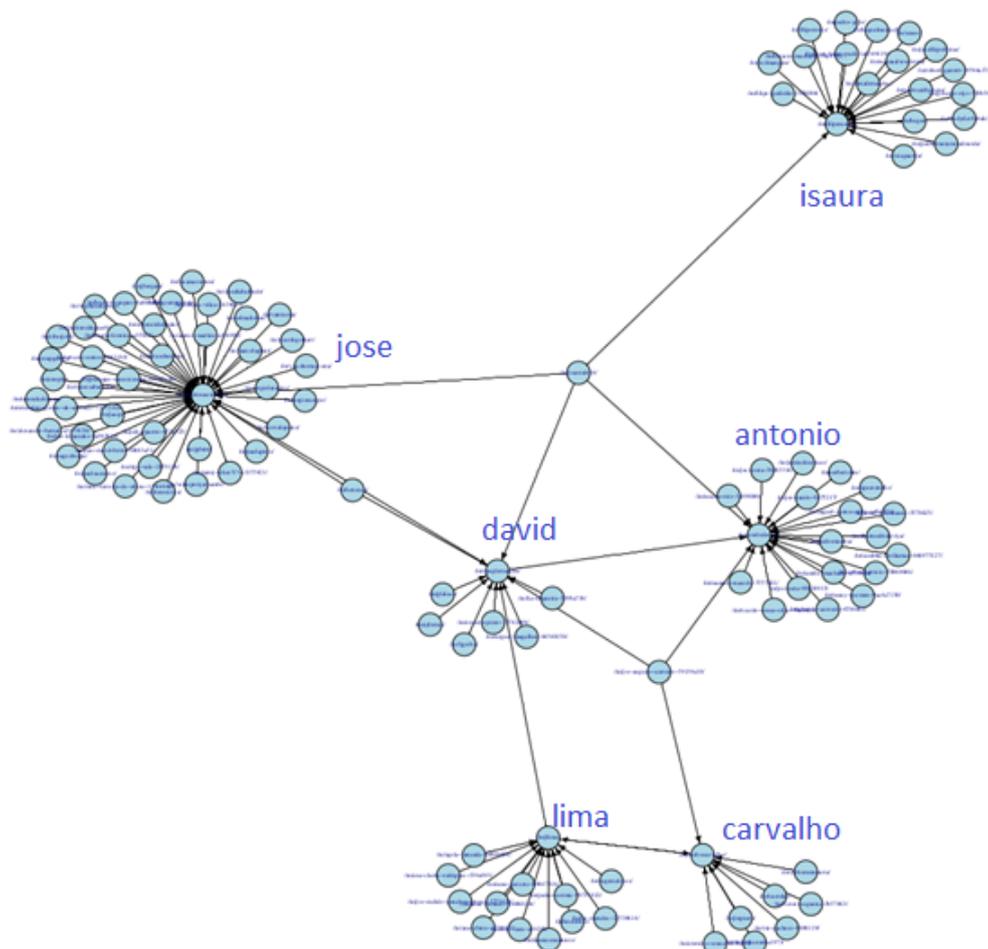


Figura 6 – Rede para a especialidade de SQL

Na Tabela 1 indica-se o valor do PageRank para os seis candidatos melhor classificados na especialidade de SQL.

Tabela 1 – Candidatos da especialidade SQL com melhor PR

Candidato	PageRank
/antonio/	0.188
/david/	0.159
/jose/	0.118
/lima/	0.086
/carvalho/	0.059
/isaura/	0.052

É de notar que o candidato ‘jose’ num vértice de grau 45, tem um PageRank inferior ao candidato ‘david’ que tem um vértice de grau 10. Ainda que estes cálculos do PageRank se tenham baseado num conjunto limitado de dados, estas imagens permitem ilustrar uma muito importante característica do algoritmo. O PageRank não considera apenas o número de recomendações diretas mas também a qualidade de cada uma delas. Deve notar-se que o candidato ‘david’ é referido por ‘lima’ e por sua vez ‘lima’ por ‘carvalho’.

O processo de definição das redes de recomendação e cálculo dos valores de PageRank repete-se para as várias especialidades.

4.2 – Avaliação multicritério dos candidatos com várias especialidades

Nesta sub-secção vamos expandir a análise a quatro especialidades escolhidas: SQL, Python, C# e Java. Utilizando o algoritmo TOPSIS pretende-se encontrar os melhores candidatos a combinação de especialidades.

A Tabela 2 apresenta os valores de PageRank de dez candidatos nas especialidades de SQL, Python, C# e Java.

Tabela 2 - Candidatos com conhecimentos de sql, python, csharp e java

Candidato	sql	python	csharp	java
/antonio/	0.188670	0.357888	0.434098	0.302628
/carvalho/	0.059168	0.001807	0.003488	0.011977
/david/	0.159362	0.408678	0.422879	0.224163
/ernesto/	0.003168	0.001807	0.003488	0.011977
/falcao/	0.003168	0.001807	0.003488	0.011977
/freitas/	0.003168	0.001807	0.003488	0.011977
/isaura/	0.052315	0.084503	0.003488	0.011977
/jose/	0.118168	0.001807	0.003488	0.011977
/lima/	0.086047	0.001807	0.003488	0.011977
/nogueira/	0.003168	0.001807	0.003488	0.011977

Aplicando o algoritmo TOPSIS para a obtenção dos quatro melhores candidatos, com pesos iguais para todas as especialidades, resultam os valores da Tabela 3.

Tabela 3 – Resultado da avaliação multicritério

Candidato	Valor
/antonio/	1.0000000
/david/	0.7754720
/jose/	0.2654553
/isaura/	0.1663549

5. Conclusões

Este trabalho apresenta uma abordagem objetiva a um problema normalmente analisado sob uma perspectiva subjetiva: a avaliação de candidatos em processos de seleção de recursos humanos baseada nas recomendações da rede social LinkedIn. Esta abordagem baseada na informação dos pares, dispensa os candidatos de eventualmente introduzirem informação menos verdadeira nos CV.

O método proposto para seleção de candidatos PageRank-TOPSIS recorre a dois algoritmos conhecidos e implementados na linguagem R. Aplicou-se com sucesso o algoritmo PageRank à análise de uma rede de recomendações profissionais e o algoritmo TOPSIS na decisão multicritério para duas ou mais especialidades necessárias a um determinado perfil profissional.

Através do método proposto é possível acelerar o processo de identificação dos melhores candidatos num conjunto de dimensão corrente, mas também aplicá-lo a um grande conjunto de dados. Acreditamos que a simplicidade do modelo de dados e a baixa complexidade das rotinas implementadas permite utilizar o método em ambientes Big Data, podendo encontrar soluções sem paralelo nos métodos tradicionais.

Para além da complexidade algorítmica, o método de duas fases apresenta grande clareza e facilidade de compreensão tanto para os decisores como para os candidatos.

Referências

1. Bakar, A. A., Ting, C.-Y. (2011). Soft skills recommendation systems for it jobs: A bayesian network approach. In 3rd Conference on Data Mining and Optimization (DMO), pp 82–87.
2. Barnes, N. G., Mattson, E. (2009). Social Media Adoption Among the 2009 Inc. 500: New Tools and New Trends. www.sncr.org/sites/default/files/socialmedia2009_0.pdf
3. Bersin, J. (2012). LinkedIn is Disrupting the Corporate Recruiting Market. <http://www.forbes.com/sites/joshbersin/2012/02/12/linkedin-is-disrupting-the-corporate-recruiting-market>.
4. Brin, S., Page, L. (1998). The Anatomy of a Large-scale Hypertextual Web Search Engine. *Comput. Netw. ISDN Syst.*, 30(1-7): pp 107–117.
5. Capiluppi, A., Serebrenik, A., Singer, L. (2013). Assessing Technical Candidates on the Social Web. *IEEE Software*, 30(1): pp 45–51.
6. Csardi, G., Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.
7. Domeniconi, G., Moro, G., Pagliarani, A., Pasini, K., Pasolini, R. (2016). Job recommendation from semantic similarity of linkedin users' skills. In Proceedings of the 5th

- International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM,, pages 270–277. INSTICC, SciTePress.
8. Fernandes, D. (2015), PageRank na seleção de recursos humanos, InforAberta 2015 – Jornadas de Informática da Universidade Aberta (Montijo).
 9. Fernandes, D. (2018), O algoritmo PageRank aplicado a redes de recomendações para seleção de recursos humanos, dissertação de Mestrado em Tecnologias e Sistemas Informáticos Web, Universidade Aberta, Portugal.
 10. Gkorou, D., Pouwelse, J., Epema, D. (2015). Trust-based Collection of Information in Distributed Reputation Networks. In Proceedings of the 30th Annual ACM Symposium on Applied Computing, SAC '15, pp 2312–2319, New York, NY, USA. ACM.
 11. Gusdorf, M., Schaefer, B., Woolever, N. A., e Leonard, S. H. (2008). Recruitment and Selection: Hiring the Right Person A two-part learning module for undergraduate students. Society for Human Resource Management.
 12. Hwang, C.-L., Yoon, K. (1981). Methods for Multiple Attribute Decision Making. pp 58–191.
 13. Kessler, R., Béchet, N., Torres-Moreno, J. M., Roche, M., e El-Bèze, M. (2009). Profilage de candidatures assisté par Relevance Feedback. TALN 2009, Senlis.
 14. Pérez-Rosés, H. and Sebé, F. (2017). Iterated endorsement deduction and ranking. Computing, 99(5): pp 431–446.
 15. Pérez-Rosés, H., Sebé, F., and Ribó, J. M. (2016). Endorsement deduction and ranking in social networks. Computer Communications, 73: pp 200 – 210. Online Social Networks.
 16. Strohmeier, S., Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, 40(7): pp 2410–2420.



David Fernandes, Licenciado em Informática e Mestre em Tecnologias e Sistemas Informáticos Web, pela Universidade Aberta, Portugal. Consultor informático, apaixonado por linguagens de programação, da RPG-II à C#, da Lisp à R, e em pesquisa, transformação, organização e análise de informação heterogénea: alfanumérica, geométrica e geográfica.



Luís Cavique, Professor Auxiliar no Departamento de Ciências e Tecnologia (DCeT), Secção de Informática, Física e Tecnologia (SIFT) e investigador no BioISI-MAS da FCUL. Licenciado em Engenharia Informática em 1988 pela FCT-UNL. Obteve o grau Mestre em Investigação Operacional e Engenharia Sistemas pelo IST-UTL em 1994. Obteve o grau de Doutor em Engenharia Sistemas pelo IST-UTL em 2002. Tem como áreas de interesse, a intersecção da Informática com a Engenharia de Sistemas designadamente as áreas de Otimização Heurística e “Data Mining”.