

Estudo dos Próximos Locais a Visitar na Rede Foursquare

Eudália Bilro

Mestre pela Universidade Aberta, dalia.bilro@gmail.com

Luís Cavique

Universidade Aberta, lcavique@uab.pt

Resumo

A comunicação nas redes sociais surgiu da necessidade que o ser humano tem em partilhar assuntos, ideias, preferências comuns criando assim laços assentes em afinidades. A constante presença dos utilizadores nas redes sociais expressando as suas opiniões sobre produtos, marcas, pessoas, gostos, ou costumes tem vindo a desencadear um grande interesse por parte de empresas e pessoas em analisar essas informações. Numa sociedade que diariamente é capaz de produzir dados em massa, é cada vez mais necessária a criação de ferramentas para a sua análise e interpretação de forma a disponibilizar todo um conjunto de informações úteis para a tomada de decisões. Neste contexto, este trabalho descreve o processo de aplicação de técnicas de Data Mining em dados extraídos da rede social *Foursquare* de forma a obter informações relevantes que auxiliem na identificação de padrões de comportamentos. Através da descoberta de padrões sequenciais, este estudo irá permitir a visualização dos dados organizados numa poli-árvore com o objetivo de estudar os próximos locais a visitar na rede *Foursquare*.

Palavras-chave: redes sociais, extração dados, padrões sequenciais, *foursquare*

Abstract

Communication in social networks has arisen from the need that human beings have to share subjects, ideas or common preferences and in this way to create bonds. The constant presence of the users of social networks expressing their opinions about products, brands, people, tastes or habits has developed a great interest by companies and researchers to analyze this information. In a society that is capable of producing mass information on a daily basis it is necessary to create tools for the analysis and interpretation of such data that will provide information useful to decision-making process. In this context, this work describes the process of applying data and graph mining techniques to data extracted from the social network *Foursquare*, in order to obtain relevant information for identifying behavior patterns. Through the discovery of sequential patterns, this study will allow us to visualize the data arranged in a poly-tree with the purpose of forecasting the next locations to be visited on the *Foursquare* network.

Keywords: social networks, data mining, sequential patterns, *foursquare*

1. Introdução

Uma rede social pode ser definida como um conjunto de pessoas ou grupos de pessoas, ligados entre si por relações de vários tipos e que partilham valores e objetivos comuns.

Atualmente, e em virtude das novas tecnologias, o acesso às redes sociais assume uma forte tendência, principalmente no que respeita às redes baseadas em serviços de localização. Geralmente, trata-se de aplicações utilizadas em *smartphones* com acesso à Internet e ao GPS, em que os utilizadores podem partilhar referências a locais, num determinado sistema. As suas funcionalidades passam por permitir aos seus utilizadores a partilha do lugar onde se encontram, ligar-se a pessoas que estão no mesmo local, visualizar comentários de utilizadores com os mesmos interesses e pesquisar locais através de informações relevantes.

Tendo em vista este contexto, este artigo focar-se-á na maior rede social baseada em geolocalização – o Foursquare, uma plataforma lançada em 2009, com mais de 10 milhões de utilizadores e perto de 1 milhão de check-ins por dia. Neste trabalho é apresentado um estudo do comportamento de padrões sequenciais através dos check-ins na aplicação, que irá permitir analisar hábitos de localização futuros dos utilizadores da aplicação.

Este artigo encontra-se dividido em cinco secções. Na secção presente, procede-se à identificação do estudo realizado e dos seus objetivos. Na segunda secção é feita uma revisão da literatura existente e na terceira secção é apresentado o *dataset* Foursquare de forma a obter uma caracterização geral sobre a arquitetura da rede. Na quarta secção é apresentada a análise e discussão dos resultados das redes e poli-árvores resultantes e por fim, na quinta secção são expostas as conclusões relativamente ao tema e objetivos definidos inicialmente.

2. Trabalho relacionado

Uma das atividades de *Data Mining* é a descoberta de padrões sequenciais. A criação de algoritmos eficientes para a descoberta do conhecimento tem crescido de forma significativa devido à existência de grandes volumes de dados com dependência temporal e decorre da necessidade iminente de extrair conhecimento dessa mesma informação, uma vez que o ser humano não é capaz de interpretar uma grande quantidade de dados [Fayyad et al., 1996]. Uma série temporal é uma sequência de observações de uma variável ao longo do tempo, tomadas em intervalos temporais regulares [Wooldridge, 2000].

O problema da descoberta de padrões sequenciais foi inicialmente discutido em [Agrawal & Srikant, 1995]. Considerando uma base de dados de sequências, onde cada sequência é uma lista de transações ordenadas pelo tempo e cada transação uma lista de itens. A descoberta de padrões sequenciais consiste em encontrar subsequências frequentes que satisfaçam os critérios mínimos especificados pelo utilizador.

Este artigo irá abordar a técnica usada na descoberta de padrões sequenciais, tendo como base uma abordagem ao algoritmo Ramex [Cavique 2007, Cavique e Coelho 2008, Cavique 2015], cuja utilização permitirá detetar esses padrões pelo conceito de poli-árvores.

A abordagem principal à descoberta de padrões entre dados com dependências temporais consiste na adaptação do algoritmo tradicional de descoberta de regras de associação (*Apriori*) ao caso particular deste tipo de dados. Alguns exemplos são os algoritmos AprioriAll [Srikant & Agrawal, 1995] e GSP (*Generalized Sequential Pattern*).

Sabendo-se que uma série temporal é uma descrição do passado, um procedimento lógico para realizar previsões será a utilização desses mesmos dados históricos. Segundo o estudo apresentado em [Cavique & Coelho, 2008], a prospeção de padrões temporais pode ser dividida em quatro diferentes abordagens: os padrões periódicos, a descoberta de sequências, os episódios frequentes e os modelos das cadeias de Markov.

O nome do algoritmo *Ramex* provém do Latim, que significa “ramo”, em português. Esta abordagem apresenta a informação em rede, pois todos os itens são levados em consideração, e permite uma visão global dos dados. O seu objetivo é criar uma sequência de árvores com várias ramificações, tantas quantas as necessárias, de forma a garantir que todos os ramos serão visitados a partir do seu vértice-raiz.

Uma árvore é um grafo orientado com um e um só caminho simples entre quaisquer dois vértices. Um subgrafo que seja uma árvore e contenha todos os vértices do grafo é designado por árvore abrangente.

A abordagem do Algoritmo Ramex [Cavique 2007, Cavique e Coelho 2008, Cavique 2015] tem duas fases. Na primeira fase, dá-se a transformação da base de dados numa rede. Os dados são ordenados e, para cada linha, um novo atributo é criado: o próximo item. Em seguida é criada uma rede, i.e., um grafo com uma fonte (ou raiz) e um nó final chamado sumidouro. Esta rede, em que os ciclos são permitidos, condensa a informação da base de dados, incorporando todas as sequências possíveis. Na rede construída, cada nó corresponde a um item e a transição representa uma sequência de um item para o próximo item. O peso de cada arco corresponde ao número de vezes que um item antecede um próximo item.

Em [Cavique & Coelho, 2008], para a geração de poli-árvores, desenvolveu-se a heurística *Back-and-Forward*, também baseada no algoritmo de Prim, cuja complexidade temporal é representada por $O(N^2)$, em que N representa o número de vértices.

3. Dataset Foursquare

O *Foursquare* é uma rede social baseada em localização, também conhecida como rede social móvel [Sutko & Silva, 2011], na qual o utilizador, por meio de um telemóvel, informa acerca do lugar onde está naquele instante.

A primeira versão do *Foursquare* foi desenvolvida em meados de 2008 por Dennis Crowley e Naveen Selvadurai e permitia ao utilizador o envio de mensagens escritas de telemóvel a todos os seus contactos, avisando da sua localização.

O *Foursquare* é definido como a ferramenta de pesquisa mais completa e com maior número de utilizadores no mundo, aproximadamente 45 milhões em termos globais, obtendo 1 milhão de registos mundiais a cada mês. Diariamente são adicionados 35 mil novos perfis e o número de *check-ins* ultrapassa 2,5 milhões [Foursquare, 2014].

Para a realização deste estudo recorreu-se a dados primários disponibilizados em <http://www-users.cs.umn.edu/~baojie/Research.htm> que serviram de base para um estudo que apresenta um sistema de recomendação baseado nas preferências pessoais de cada utilizador [Bao, Zheng & Mohamed, 2012].

A tabela *Tips & Check-in do dataset*, assume um papel bastante relevante neste estudo pois reúne os comentários de todos os utilizadores referentes aos locais existentes na aplicação. É através desta tabela que serão extraídos os *check-ins* da aplicação, identificando o ID do local, o ID do utilizador, a data e o comentário em relação a esse local.

A tabela 1 mostra o número utilizadores, lugares, *check-in* e categorias do *dataset Foursquare*.

Tabela 1 – Dataset Foursquare

Nº. utilizadores	Nº. lugares	Nº. <i>check-ins</i>	Nº. Categorias
49.001	157.655	315.852	9

Na tabela 2 podemos observar os rácios das variáveis *UserId* e *VenueId* do *dataset*. Deve-se ressaltar que o número médio de *check-ins/utilizador* é bastante inferior ao apresentado em outros estudos [Cheng et al., 2011].

O rácio de *check-in* por local é de 3,22 e o rácio de *check-in* por utilizador é de 3,79. Os valores apresentados representam rácios baixos o que poderá indicar que os utilizadores não são utilizadores frequentes da aplicação e que o número de locais visitados por esses utilizadores também não é elevado. Na prática poderá também indicar um baixo número de sequências encontradas, o que levará de certa forma a apresentar uma alternativa para a redução de dados em subconjuntos com vista a estudar diferentes abordagens.

Tabela 2 - Rácios das variáveis *UserId* e *VenueId*

Rácios	Utilizadores 48.308	Locais 56.854	Nº. <i>check-ins</i> 183.211
Utilizadores	---		
Locais	1,17	---	
Nº. <i>check-ins</i>	3,79	3,22	---

Durante a fase de compreensão dos dados que incluiu a identificação da informação relevante para o estudo, foi possível concluir que: quanto ao perfil dos utilizadores deste *dataset Foursquare*, maioritariamente, não são utilizadores frequentes da aplicação, pois existe uma baixa percentagem com mais de 6 *check-ins* efetuados. Os locais mais visitados encontram-se distribuídos essencialmente por 3 zonas e a grande maioria foi visitada, pelo menos, mais de 2 vezes.

Os utilizadores recorrem à aplicação numa vertente mais social, já que a categoria mais visitada é *Food* e *NigthLife* - nos dias da semana/horas conforme a semana/hora vai avançando, existe um aumento de atividade. Importa destacar, também, os baixos valores encontrados no que respeita aos rácios de *check-in* por local e por utilizador.

4. Análise dos resultados

4.1. Visualização da rede utilizando k-cores

As estruturas de grafos muito grandes podem ser estudadas a partir da segmentação em partes menores, que são mais simples de serem manipuladas.

Uma decomposição possível é baseada em *k-cores onde o subgrafo induzido H_k é um k -core*, ou núcleo de ordem k , se todo vértice de H_k possui grau menor ou igual à k e H_k for o maior grafo com esta propriedade. O *core* de maior ordem é chamado de *core principal* e o número de *core* de um vértice, v , é a menor ordem de core que contém este vértice, denotado por $k_S(v)$.

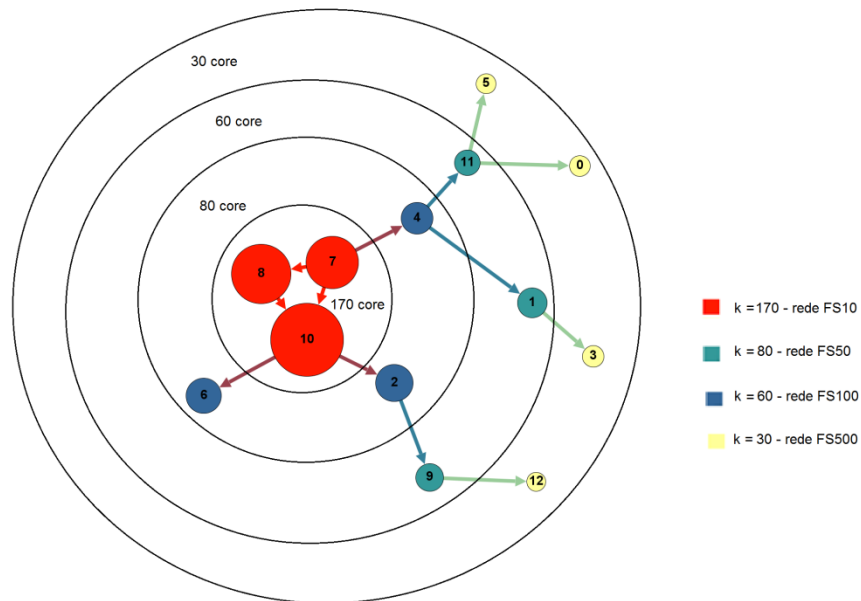


Figura 1 - Rede original com a divisão em núcleos

A metodologia utilizada nesta análise baseou-se em um estudo da rede que realiza a decomposição em *k*-cores baseado no número de *check-ins*. É baseado na remoção dos nós de acordo com o número de *check-ins*, onde *k* indica o número de *check-ins* mínimo que todo elemento do núcleo possui. Uma Rede *k*-core significa que todos os locais têm visitas mínimas igual a *k*. A figura 1 exemplifica o método de decomposição em núcleos.

Na primeira etapa exclui-se aqueles vértices com *k* inferior ou igual a 30. Os vértices que não foram removidos, bem como as arestas que conectam os mesmos formam o núcleo $k_S = 30$, ou seja, a rede FS500. A seguir, repete-se o mesmo procedimento para os demais locais presentes na rede para *k* inferior a 60, 80 e 170 respetivamente para as redes FS100, FS50 e FS10.

4.2. Caracterização da Rede

Foram criadas redes de diferentes dimensões que permitissem um estudo detalhado da rede. As estruturas resultantes dessa decomposição foram estudadas a partir do Software Gephi [Gephi, 2015] que permitiu obter as medidas de análise das diferentes redes e identificar os locais mais destacados de cada rede. Foram identificados os nós com maior centralidade utilizando as métricas disponíveis no Gephi em particular as medidas Hit.

As redes formadas foram analisadas de acordo com os seguintes parâmetros:

Tabela 3 - Resultado dos parâmetros de cada rede

Locais	Utilizadores	Check-In	Check-In/Locais	Nome Rede
56.854	48.308	183.211	3,22	FS
500	15.980	26.671	53,34	FS500
100	8.087	10.732	107,32	FS100
50	5.809	7.209	144,18	FS50
10	2.424	2.792	279,2	FS10

Numa primeira análise à tabela 3 nota-se que, a rede original (FS) contém um elevado número de locais e um rácio *check-in*/local muito baixo. Esse valor vai crescendo conforme a diminuição do número de locais, sendo que, a rede FS100, FS50 e FS10 apresentam já valores mais fáceis de serem analisados.

Com vista a identificar os nós mais relevantes na rede existiu a necessidade de recorrer a uma serie de estatísticas capazes de efetuar uma análise cuidada das redes. As medidas de redes são fundamentais para a sua caracterização, representação, classificação e análise.

A tabela 4 resume as medidas estatísticas utilizadas para este estudo e que irão ser fundamentais na interpretação de cada uma das redes.

Tabela 4 - Resultado das medidas Referentes às Redes analisadas

Nome da Rede	Densidade	Grau Médio Ponderado	Diâmetro Rede	Comprimento Caminho Médio	Caminhos curtos
FS500	0,01	1.868,00	6	3,06	244.533
FS100	0,16	28,25	4	1,95	9.801
FS50	0,28	28,00	3	1,79	2.450
FS10	0,80	36,80	2	1,33	90

Das análises efetuadas às diferentes redes é verificável que para além de as redes serem pouco dispersas, existem alguns locais com relevância que se destacam. Se por um lado maior parte das redes apresentam uma baixa densidade por outro há entidades que nesta lógica de precariedade de relacionamentos já referidos, acabam mesmo assim por ocupar uma posição aparentemente privilegiada no quadro geral da rede.

Pelo número de caminhos curtos existentes em todas as redes pode-se afirmar que não existem sequências longas, os utilizadores dispersam-se pelos múltiplos locais e mesmo entre pares de locais não existem muitas passagens.

Apesar de encontrar os nós de maior centralidade a densidade gráfica das redes não permitem identificar os percursos dos utilizadores com clareza.

4.3. Poli-árvores

O algoritmo Ramex é composto por duas fases, a fase de transformação do *dataset Foursquare* numa rede cíclica e uma segunda fase de procura das sequências frequentes representadas numa poli-árvore. Com recurso ao algoritmo Ramex foi possível identificar as poli-árvores dos percursos dos utilizadores da rede *FourSquare*.

Poli-árvore FS10

Na poli-árvore resultante da rede FS10 (figura 2) há a destacar que as sequências mais pesadas são compostas pelos nós (4,7,9) e (5,3,2) com peso de 14, os quais estão devidamente assinalados a laranja. As sequências mais pesadas são constituídas apenas por 1 categoria, ou seja, todos os nós dentro dessa sequência pertencem à mesma categoria. A árvore ponderada de maior peso (árvore que visita todos os nós) tem um comprimento total igual a 51 (8 + 7 + 6 + 7 + 4 + 3 + 11 + 3 + 2).

Poli-árvore 50

Na poli-árvore da rede FS50 (figura 3), foi possível identificar os pontos da rede FS10, mostrando que todos eles se encontram em posições de destaque e mostrando que esta rede continua a pertencer ao núcleo da rede FS50.

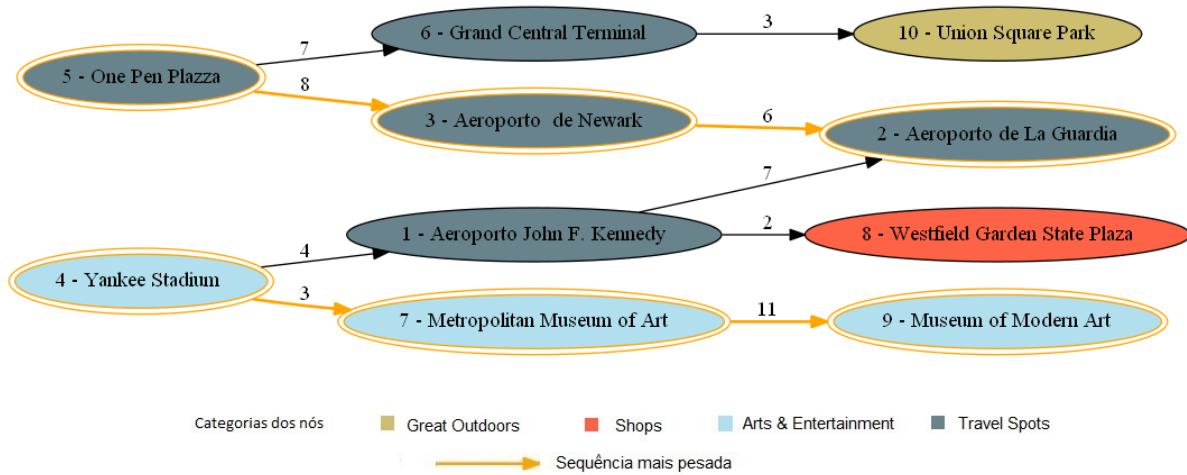


Figura 2 – Locais e Categorias da Rede FS10

Os pesos das arestas não são muito elevados, mas as relações encontradas foram significativas mostrando que, mesmo para locais que não estão fortemente relacionados, ligações relevantes ainda são encontrados usando este algoritmo.

Embora os maiores pesos estejam diretamente relacionados com os nós pertencentes à rede FS10, os pesos entre os nós são distribuídos equitativamente, não se notando grandes diferenças de valores entre categorias e zonas da poli-árvore.

Tal como a poli-árvore resultante da rede FS10, os nós com categorias iguais tendem a estar próximos uns dos outros, dividindo a poli-árvore em dois grandes grupos. O primeiro grupo mais localizado no topo da poli-árvore, contém uma grande parte de locais da categoria *Food* e a zona mais abaixo as restantes categorias, *Travel Spots*, *Arts & Entertainment*, *Great Outdoors* e *Shops*.

Neste contexto, pode-se dizer que a zona superior por ser a zona com mais desmultiplicação de nós, pode ser a zona de utilizadores frequentes que usam a aplicação de forma mais social. Enquanto que, a zona abaixo é a zona de utilizadores mais esporádicos da aplicação, ou seja, contém nós que originam sequencias menos profundas onde os utilizadores utilizam a aplicação de forma mais ocasional e numa vertente mais turística.

A árvore ponderada de maior peso (árvore que visita todos os nós) tem um comprimento total igual a 211 para um total de 50 nós (locais) e 690 vértices (ligações entre locais).

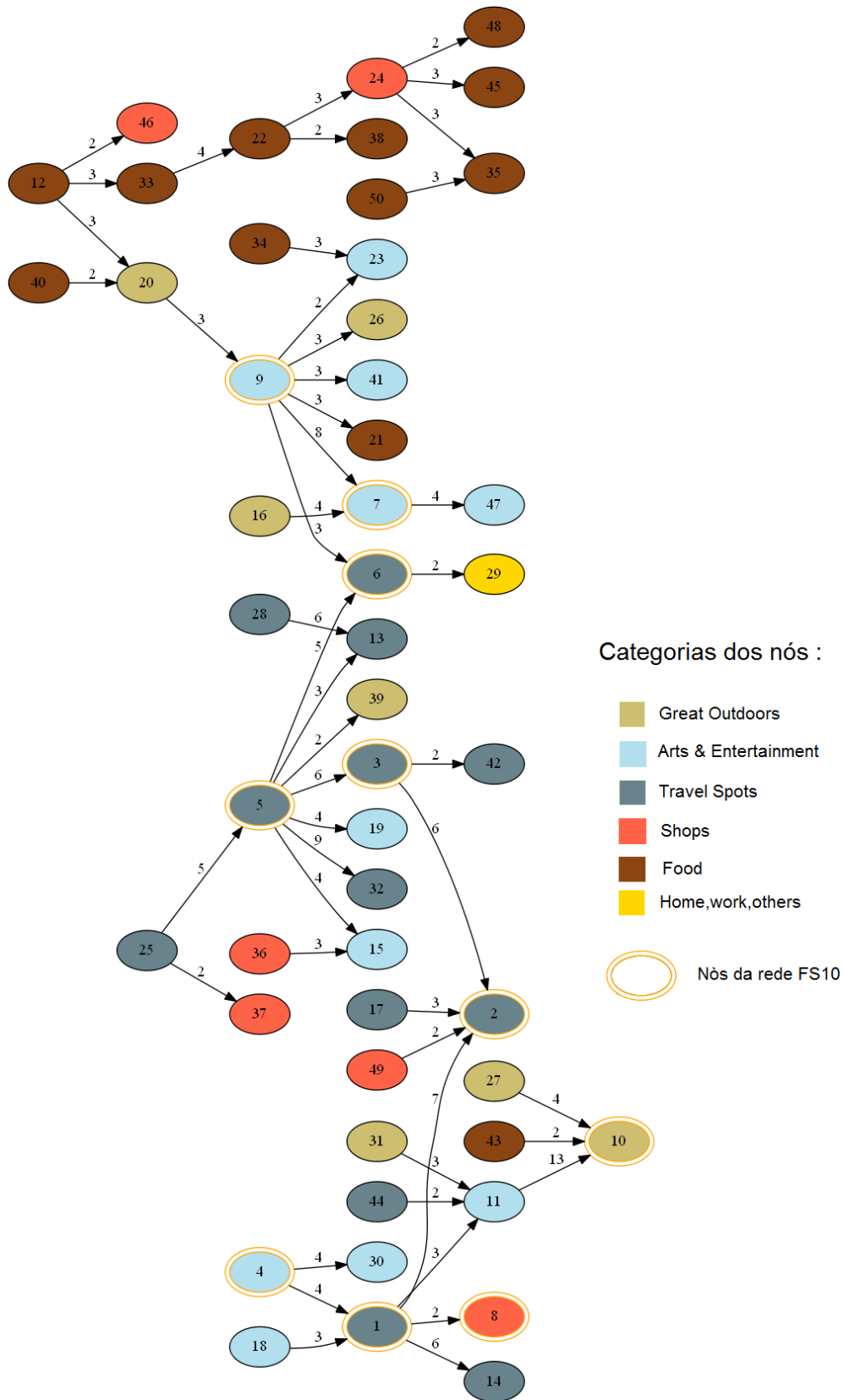


Figura 3 - Locais e Categorias da Rede FS50

Poli-árvore da rede FS100

Na poli-árvore da rede FS100 (figura 4), conseguiu identificar-se os pontos da rede FS50, mostrando que todos eles se encontram em posições de destaque e mostrando que esta rede continua a pertencer ao núcleo da rede FS50. Os pesos das arestas continuam sem apresentar valores muito elevados o que permite afirmar que não existem sequências com valores muito pesados, ou seja, que os utilizadores da aplicação dispersam-se um pouco entre a transição entre locais.

Mais uma vez se constata que os maiores pesos, embora não sejam valores muito relevantes em relação aos restantes, continuam de certa forma relacionados com nós pertencentes à rede FS10.

A tendência existente para a divisão da poli-árvore em dois grandes grupos continuou a verificar-se. A predominância de ligações com mais profundidade na parte de cima da poli-árvore é onde os locais da categoria *Food* têm maior destaque.

Podíamos esperar que os utilizadores utilizassem racionalmente a rede *FourSquare* para identificar todos os pontos onde passam, gerando uma reportagem dos seus circuitos. Ora, esta presunção não existe na sua totalidade. Existem utilizadores que registam sequências de locais de categoria *Food* e outros que registam consecutivos locais de categoria *Travel*.

Poli-árvore da rede FS500

Devido à impossibilidade de apresentar graficamente o resultado da poli-árvore da rede FS500, optou-se por fazer um pequeno estudo das sequências geradas de forma a completar o estudo de todas as redes.

Quanto às sequências detetadas, verificou-se que todas elas apresentam uma percentagem relativa bastante baixa o que leva a concluir que cada utilizador procura os seus próximos locais de interesse, afastando-se dos restantes.

Quanto ao comportamento dos utilizadores agrupados por categoria, consegue-se perceber uma clara tendência para visitas entre locais da categoria (*Food* -> *Food*) o que reforça a ideia de que esses utilizadores utilizam a aplicação diariamente e de forma mais social, partilhando para os seus contactos a sua localização e as suas preferências.

Seria expetável que os utilizadores da rede *Foursquare* fizessem uso da aplicação da sua totalidade, no entanto tal situação não se verifica. Através da análise detetou-se que existem utilizadores a usarem a aplicação para registar a sua atividade em duas grandes vertentes: registo de sequências em locais da categoria *Food* e locais relacionados com a categoria *Travel*.

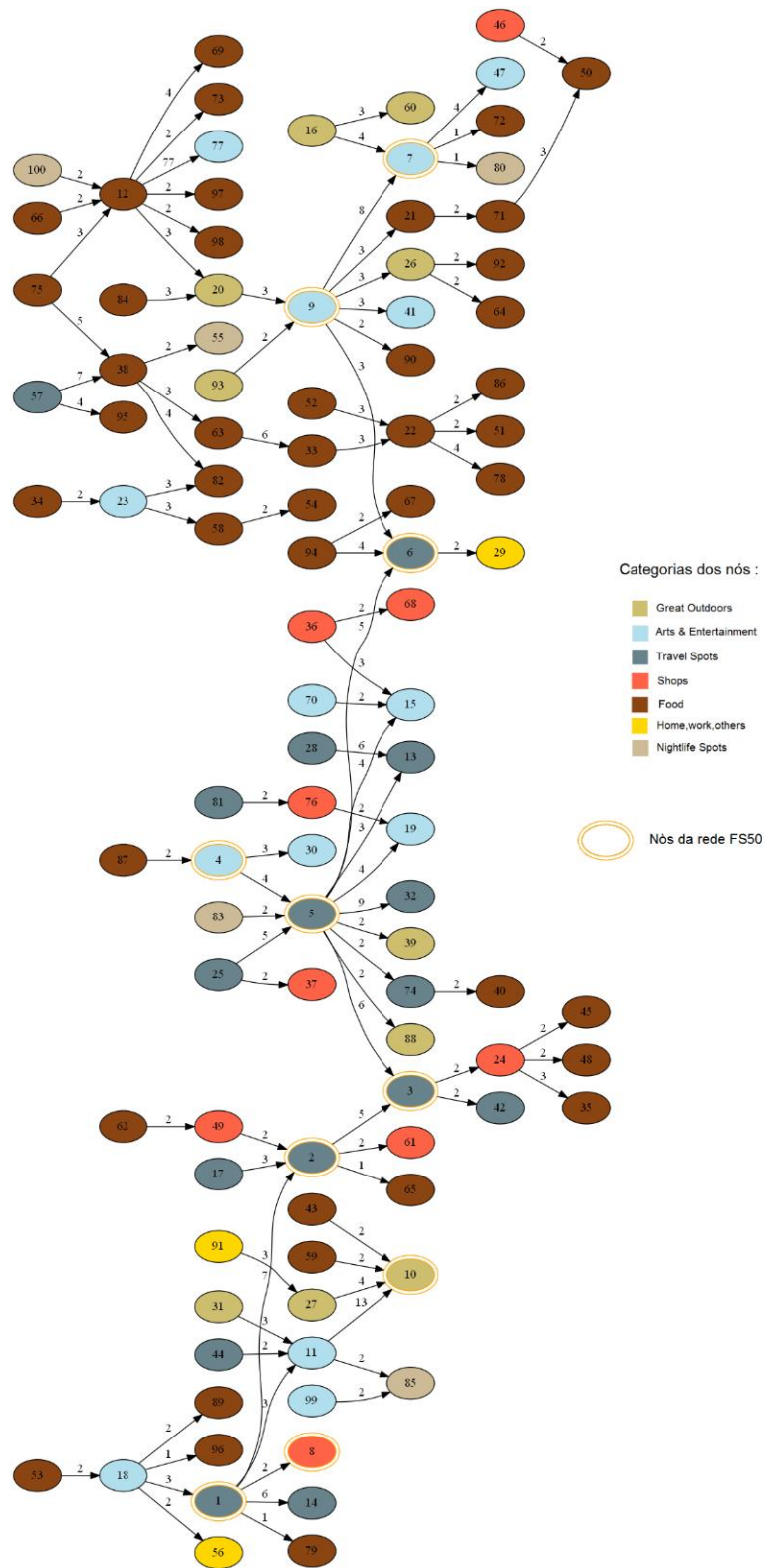


Figura 4 – Locais e Categorias da Rede FS100

5. Conclusão

Este trabalho apresentou uma nova abordagem ao uso de algoritmos que permitam uma visão global dos dados de uma rede possibilitando a visualização dos dados organizados numa poli-árvore de forma a obter sequências frequentes. Uma das principais inquietações iniciais deste estudo foi sem dúvida o grande volume dos dados que o *dataset* apresentava.

Após inúmeras experiências com os dados pré-processados, verificou-se que a divisão do *dataset* em núcleos mais pequenos originando redes de diferentes dimensões iria contribuir de forma significativa para o objetivo deste estudo. Sobre estas diferentes redes foi aplicado o algoritmo Ramex [Cavique 2007, Cavique e Coelho 2008, Cavique 2015] e os resultados obtidos levam a acreditar no sucesso da aplicabilidade do modelo em situações práticas que requerem o estudo do comportamento humano.

Baseado nessa análise visual e estatística, concluiu-se que existe uma grande dispersão dos percursos dos utilizadores, fazendo com que não existam longas sequências e que a transição entre locais seja baixa. Os locais mais relevantes pertencem à categoria *Travel Spots* e são comuns a maior parte das redes estudadas.

Os resultados obtidos na fase de geração de poli-árvores pelo algoritmo Ramex apontam sobretudo para uma rede onde os utilizadores procuram os seus próximos locais de interesse e que se dispersam pelos diferentes locais na rede criando percursos bastante distintos conforme a frequência com que utilizam a aplicação.

A aplicação do Algoritmo Ramex à rede *Foursquare* obteve bons resultados essencialmente no que respeita à descoberta de sequências frequentes na poli-árvore de maior peso, objetivo subjacente a este trabalho. As sequências descobertas apontam para uma certa tendência onde os utilizadores procuram os seus próximos locais de interesse afastando-se dos restantes, criando diferentes percursos.

Embora as sequências descobertas não tenham resultado em valores absolutos muito altos, as ligações entre locais foram devidamente analisadas apresentando possíveis justificações para a sua relação.

Detetou-se também através da análise, que os utilizadores usam a aplicação para registar a sua atividade em duas grandes perspetivas: a perspetiva de utilizador frequente que regista sequências em locais da categoria *Food* e a perspetiva de utilizador ocasional que efetua *check-ins* em locais relacionados com a categoria *Travel*. Desta forma, exclui-se a hipótese de que os utilizadores da rede *Foursquare* utilizam a aplicação na sua totalidade e que cada utilizador cria os seus próprios percursos em determinadas categorias.

Apesar da satisfação global com os resultados obtidos, uma das melhorias sugeridas em trabalhos futuros, diz respeito à possibilidade de gerar a Árvore Geradora Mínima para a rede *Foursquare*, de forma a criar um registo dos locais menos visitados da rede. Uma área de interesse para empresas de estudos de mercado e/ou para campanhas de Marketing empresarial de forma a inverter uma tendência menos favorável em visitas de um determinado local.

Referências

- R. Agrawal, R.Srikant - Mining Sequential Patterns. Proceedings 11th International Conference Data Engineering, ICDE, pp. 3–14, IEEE Press,1995.
- J. Bao, Y. Zheng, M. Mokbel - Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data. In ACM SIGSPATIAL (GIS 2012), pp.199-208, Redondo Beach, CA, US, 2012.
- L. Cavique - A network algorithm to discover sequential patterns. In Proceedings of the Artificial intelligence 13th Portuguese conference on Progress in artificial intelligence, EPIA'07, LNAI
- L.Cavique, J.Coelho - Descoberta de Padrões Sequenciais utilizando Arvores Orientadas. Revista de Ciências da Computação, volume III, ano III, nº3, pp. 12-22, 2008.
- L.Cavique - Ramex: A Sequence Mining Algorithm Using Poly-trees. New Contributions in Information Systems and Technologies, Advances in Intelligent Systems and Computing, A.M Rocha, S. Correia, L.P Costanzo, Eds Reis, Springer edition, volume 2, pp. 143-154, 2015.
- Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui - Exploring Millions of Footprints in Location Sharing Services. In Proceedings of the Fifth International Conference on Weblogs and Social Media, ICWSM, pp. 81–88, 2011.
- Blog do Foursquare, <http://blog.foursquare.com/> [Setembro de 2014]
- U.Fayyad, G.Piatetsky-Shapiro, P. Smyth, R. Uthurusamy - Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence Menlo Park, USA, 1996.
- Gepji Página oficial <http://gephi.github.io/> [Fevereiro 2015]
- R. Srikant, R. Agrawal - Mining sequential patterns: Generalizations and performance improvements. Proceedings 5th International Conference Extending Database Technology, EDBT, 1057, pp. 3-17, 1996.
- D. Sutko, A. de Souza E Silva - Location-aware mobile media and urban sociability. New Media & Society, v. 13, pp. 807-823, 2011.
- J. Wooldridge - Introductory Econometrics: a Modern Approach. SouthWestern College Publishing, a division of Thomson Learning, pp.191-193, 2000.



Eudália da Conceição Paraneta Billo, frequenta o mestrado de Tecnologias e Sistemas Informáticos Web da Universidade Aberta e tem como principais interesses de investigação as áreas de Data Mining e Redes Sociais. Foi professora do ensino secundário e formadora de Tecnologias de Informação, atualmente é consultora de Software na área de Gestão.



Luís Cavique, Professor Auxiliar no Departamento de Ciências e Tecnologia (DCeT), Secção de Informática, Física e Tecnologia (SIFT). Coordenador da Licenciatura em Informática no biénio 2012 - 2014. Licenciado em Engenharia Informática em 1988 pela FCT-UNL. Obteve o grau Mestre em Investigação Operacional e Eng. Sistemas pelo IST-UTL em 1994. Obteve o grau de Doutor em Eng. Sistemas pelo IST-UTL em 2002. Tem como áreas de interesse, a intersecção da Informática (Computer Science) com a Engenharia de Sistemas (Management Science) designadamente a área de “Data and Graph Mining”.