

IDENTIFICAÇÃO DE FRONTEIRAS DE CARREIRAS USANDO DE COMPRESSÃO DE FLUXO EM GRAFOS

RONIE MIGUEL ULIANA*, LEANDRO NUNES DE CASTRO*

* *Universidade Presbiteriana Mackenzie*
Rua da Consolação, 930
CEP 01302-907
São Paulo, SP, Brazil

Emails: ronie.uliana@gmail.com, lnunes@mackenzie.br

Abstract— Understanding how the career evolves is important to the individual, so he or she can make better decisions; to the companies, as they can plan better their internal organization; and to the governments, so they can plan better their interventions on the labor market. Here we use community detection techniques on a real database with millions of professional backgrounds to objectively identify career boundaries in Brazil and to study their topologies. Our results provide a quantitative basis for career models.

Keywords— Network, Flow Compression, Community Detection, Occupations, Boundaryless Career, Protean Career, Occupational Pole, Occupational Island.

Resumo— Compreender como a carreira profissional evolui é importante para o indivíduo, pois ele pode tomar melhores decisões de trabalho; para as empresas, pois podem planejar melhor sua organização interna; e para governos, pois podem planejar melhor suas intervenções no mercado de trabalho. Nesse artigo, técnicas de detecção de comunidades em rede foram utilizadas em um banco de dados real com milhões de experiências profissionais para identificar objetivamente as fronteiras de carreiras no Brasil e para estudar sua topologia. Os resultados fornecem base quantitativa para modelos de carreiras e novas terminologias são propostas.

Palavras-chave— Redes, Compressão de Fluxo, Detecção de Comunidades, Ocupações, Carreira sem Fronteiras, Carreira Proteana, Polo Ocupacional, Ilha Ocupacional.

1 Introdução

Cada trajetória profissional é bastante particular. Enquanto alguns seguem caminhos bem definidos, como inúmeros presidentes de empresas que começaram como estagiários, outros trilham por sequências improváveis de ocupações, como o início da carreira de Sílvio Santos, que foi Camelô e Paraquedista Militar antes de começar a carreira como Apresentador de TV (Morgado, 2017); ou Monja Coen, que foi jornalista até seus 36 anos antes de se tornar uma Monja Budista (Steiner, 2009).

Os modelos de carreiras também têm se modificado e teorias como as das *Carreiras Proteanas* e *Carreiras sem Fronteiras* advogam um distanciamento das carreiras comuns em favor de trajetórias com foco maior no indivíduo (Bendassolli, 2009), com uma movimentação mais livre entre ocupações.

No entanto, existem dificuldades naturais na transição entre ocupações e quando ela é percebida por um número suficiente de pessoas, surge uma *fronteira de carreira* (Gunz et al., 2007). Essas fronteiras delimitam um grupo de ocupações nas quais um indivíduo tem maiores chances de permanecer durante sua trajetória profissional, denominado aqui como *ilha ocupacional*.

Esse trabalho utiliza o conceito de fronteira de carreira (Gunz et al., 2007), um banco de dados real com milhões de experiências profissionais (VAGAS Tecnologia, 2014) e técnicas de detecção de comunidades em redes (Rosvall et al., 2009; Edler

et al., 2017) para encontrar ilhas ocupacionais no mercado de trabalho brasileiro, discutir e entender sua estrutura. O estudo revela onde estão as fronteiras das ilhas, quais suas composições e apresenta *insights* sobre suas estruturas.

A principal contribuição desse trabalho está em tornar concreto o conceito de *fronteiras de carreira*, aplicando técnicas de Ciência de Redes para revelar a estrutura da movimentação profissional no Brasil. O trabalho se estende nesse processo, derivando o conceito de ilha ocupacional a partir das fronteiras de carreiras como grupos de ocupações que se isolam de outros, onde a movimentação inter-ilhas é rara em comparação à movimentação intra-ilhas.

Explorando esse conceito, o trabalho caracteriza a topologia das ilhas ocupacionais e propõe a identificação de *polos ocupacionais* como ocupações que possuem papel crucial na estrutura. A topologia encontrada na maior parte das ilhas tende a um formato estrelado, sugerindo que a atuação nos polos afeta a ilha como um todo.

Esse artigo está organizado da seguinte forma. A Seção 2 descreve conceitos sobre carreira, fronteira de carreira, ilhas ocupacionais, polos ocupacionais e outros conceitos associados relevantes, enquanto a Seção 3 descreve o banco de dados utilizado na pesquisa. As Seções 4 e 5 descrevem os algoritmos empregados para detecção de comunidades e a medição utilizada na caracterização da topologia. Finalmente, a Seção 6 discute e levanta questões sobre os resultados obtidos.

2 Sobre Carreira e Ocupações

Segundo (Arthur et al., 1989), carreira é “uma sequência evolutiva da experiência profissional de uma pessoa no tempo”¹. Apesar das discussões sobre as mudanças nos modelos de carreira, passando de um modelo linear e centrado na organização para um modelo centrado no indivíduo, as definições de carreira são frequentemente associadas à progressão profissional do indivíduo (Baruch, 2004; Sullivan and Baruch, 2009; Bendassolli, 2009).

Para os fins desse trabalho, define-se carreira como a *sequência de ocupações pela qual um indivíduo passa em sua vida profissional*. Essa definição é similar à de (Arthur et al., 1989), porém, permite uma análise menos subjetiva, uma vez que ocupações profissionais podem ser extraídas de currículos e analisadas quantitativamente. No entanto, ela se torna mais limitada, já que essa definição exclui aspectos psicológicos ou sociais.

Essa pesquisa empresta o conceito de *fronteiras de carreira* (*career boundaries*) descrito por (Gunz et al., 2007) para dar significado ao trabalho. A fronteira de carreira significa que uma mudança entre ocupações nem sempre pode ser realizada livremente.

As barreiras não se limitam ao conhecimento, quaisquer dificuldades na movimentação podem criar fronteiras. Por exemplo, alguém morando em um grande centro urbano dificilmente exerceria a ocupação de “Agricultor” sem mover-se para o campo. Um “Diretor Financeiro” precisaria adequar seu padrão de vida antes de uma transição para uma ocupação com ganhos mais modestos. Uma profissão que está desaparecendo, como “Contínuo”, possui barreiras mais altas do que uma nascendo, como “Analista de Experiência do Usuário”.

Um dos pontos principais desse trabalho está na argumentação de (Gunz et al., 2007) sobre como uma fronteira subjetiva se torna objetiva. Em sua argumentação as fronteiras de carreira são subjetivas e pessoais, onde cada um tem para si quais transições podem ser feitas em sua própria carreira. No entanto, elas se tornam objetivas quando um número suficientemente grande de pessoas possui a mesma compreensão sobre essas transições, a ponto dela ser perceptível em um nível macroscópico. Dessa maneira, as fronteiras de carreira são definidas objetivamente quando uma quantidade suficiente de pessoas entra em *consenso* sobre quais são as transições incomuns.

Partindo desse pressuposto, a identificação de trajetórias comuns e incomuns é condição suficiente e necessária para a detecção de fronteiras entre carreiras. Suficiente, pois a própria definição de fronteira é dependente da identificação do que

¹No original: “an evolving sequence of person’s work experience over time”.

são transições “incomuns”. Necessária, pois não existem fronteiras objetivamente definidas sem o consenso.

Essas fronteiras isolam algumas ocupações das outras, criando um grupo coeso pelo distanciamento dos outros grupos e não por alguma característica intrínseca. Ou seja, a fronteira define o grupo, e não o contrário (Gunz et al., 2007; Abbott, 1995).

No presente trabalho esses aglomerados de ocupações são chamados *ilhas de ocupações* ou *ilhas ocupacionais*, refletindo a facilidade de movimentação dentro de suas fronteiras e o distanciamento de outros grupos.

3 O Mapa de Carreiras

O Mapa VAGAS de Carreiras (VAGAS Tecnologia, 2014) é uma rede que resume as transições de profissionais entre ocupações no mercado de trabalho. Nele, cada nó representa uma ocupação e as conexões representam o número de profissionais que se movimentou entre elas, ou seja, nas suas carreiras deixaram a ocupação anterior e passaram a trabalhar em uma nova.

É possível observar parte do Mapa VAGAS de Carreiras (MCar) na Figura 1 com as ocupações relacionadas à profissão de “mídia”. Nela, por exemplo, 16 pessoas passaram de “supervisor-de-mídia” para a ocupação “gerente-de-mídia” em sua trajetória profissional.

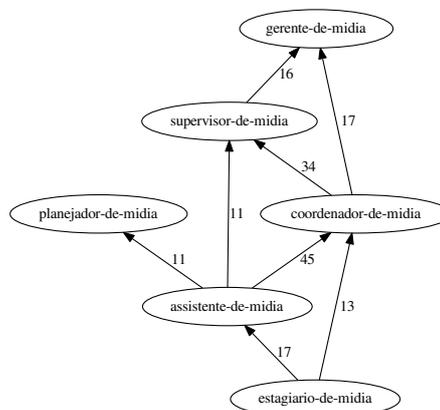


Figura 1: Parte do MCar com ocupações relacionadas à carreira de mídia.

As ocupações no MCar foram definidas a partir de *consenso* nos currículos. Como o título da ocupação é um campo em que o usuário digita livremente, assumiu-se que, se um grupo “suficientemente grande” de pessoas entra em acordo sobre uma certa nomenclatura, ela representa objetivamente uma ocupação. Essa abordagem é colocada de maneira implícita em (Abbott, 1995), enquanto

(Gunz et al., 2007) a exploram para definir fronteiras de carreiras.

Uma das maiores dificuldades para se estabelecer uma ocupação através de consenso está em definir o quanto ele deve ser “suficientemente grande” para que seja considerado como tal. Para os fins desse trabalho, esse número foi obtido de forma experimental e empírica. Para isso, procurou-se o número mínimo de repetições na grafia das ocupações de maneira que não fossem encontrados erros de digitação recorrentes, ou seja, que das ocupações criadas por consenso, nenhuma fosse mero resultado de erro. Esse número foi definido em 30 repetições da mesma grafia no título da ocupação.

Para a criação do MCar foram usados os currículos anonimizados de 10 milhões de usuários registrados no site VAGAS.com.br. A seleção de dados, bem como o processamento das informações, seguiu um procedimento conservador. Isso significa que as decisões tomadas em sua construção procuram minimizar os erros decorrentes da qualidade dos dados, mesmo que isso signifique trabalhar com uma quantidade menor deles. Utilizando o exemplo na Figura 1, isso significa que, provavelmente, muito mais do que 16 pessoas na base de dados fizeram a transição de “supervisor-de-midia” para a ocupação “gerente-de-midia”, porém, esses 16 são livres de erros e subjetividade.

Da massa de currículos, apenas os atualizados no período entre 2011 e 2016 foram usados. Currículos em duplicidade foram removidos usando o CPF como identificador; em caso de duplicidade, apenas o currículo mais recente foi considerado. Na impossibilidade de se verificar a duplicidade, como por exemplo, na ausência do CPF, o currículo também foi descartado.

Finalmente, algumas informações gerais foram extraídas dos currículos que passaram pelo processo acima e o restante das informações foi descartada, incluindo quaisquer maneiras de se identificar a pessoa descrita no currículo, garantindo a anonimidade do processo.

Os movimentação entre ocupações de cada currículo são agrupadas para formar as conexões da rede final com 7.267 ocupações e 73.064 conexões. A rede foi então armazenada em um banco de dados de grafo. Um sistema online² disponibiliza as informações publicamente. Apesar de gratuito e de consulta pública, os dados do Mapa VAGAS da Carreira não são de uso livre. A empresa gentilmente cedeu os dados ao pesquisador para esse trabalho.

4 Grafos e Detecção de Comunidades

Para esse trabalho, a identificação de comunidades por fluxo é adequada por concepção. Como exposto na Seção 2, as fronteiras de carreira são

²Disponível em <http://www.vagas.com.br/mapa-de-carreiras>

definidas pelo consenso das movimentações profissionais: transições menos frequentes definem essas fronteiras.

O algoritmo Infomap (Edler et al., 2017; Rosvall et al., 2009) utiliza *random walkers* para identificar comunidades. Os caminhos em que eles circulam mais frequentemente são considerados como fazendo parte da mesma comunidade; por outro lado, caminhos raramente utilizados definem as fronteiras entre uma comunidade e outra, de maneira análoga à concepção de fronteiras de carreira de (Gunz et al., 2007).

Portanto, por analogia, a identificação de comunidades do algoritmo Infomap identifica fronteiras de carreiras em uma rede em que as conexões representam o fluxo de profissionais entre ocupações, como é o caso do MCar.

Segundo (Grünwald, 2007), quaisquer regularidades em um conjunto de dados podem ser usadas para comprimi-lo, ou seja, descrevê-lo usando uma quantidade menor de símbolos. Quanto maior a regularidade, maior a compressão, dessa forma, dentre várias representações possíveis dos dados, a que apresentar maior compressão é também aquela que melhor identifica padrões nos dados. A Descrição de Comprimento Mínimo (*Minimum Description Length*) é a disciplina que estuda essa relação (Grünwald, 2007).

O algoritmo Infomap aproveita essa dualidade entre detecção de padrões e compressão de dados para identificar padrões de fluxo em redes. Para isso, calcula-se a probabilidade estacionária de cada nó usando o algoritmo proposto por (Page et al., 1999) aplicada a equação de entropia de Shannon (Shannon, 1948) para se obter a média do menor número de bits necessário para representar um caminho ergódico (infinito).

Assumindo que os códigos podem ser reaproveitados entre comunidades desde que existam códigos específicos para identificar saída e entrada em outras comunidades, o melhor particionamento possível passa a ser aquele que minimiza o número de bits necessário para representar caminhos ergódicos na rede segundo esse particionamento. A equação que fornece esse valor é a Equação de Mapa (*Map Equation*) (Rosvall et al., 2009).

A Equação de Mapa é

$$L(M) = q_{\rightarrow} H(Q) + \sum_{i \in M} p_{\rightarrow}^i H(P^i), \quad (1)$$

onde $L(\cdot)$ é o menor tamanho de codificação possível para representar um certo particionamento; M é o particionamento dos nós em comunidades $i \in M$; q_{\rightarrow} é a probabilidade de um *random walker* sair de qualquer comunidade; $H(Q)$ (*eta* em maiúsculo) é o número médio de bits usado para representar a entrada em uma nova comunidade segundo a distribuição de probabilidade Q ; e $H(P^i)$ também é o número médio de bits, mas dessa vez daqueles usados na comunidade i segundo a distribuição

\mathcal{P}^i ; finalmente $p_{\mathcal{O}}^i$ é a probabilidade do *walker* permanecer na comunidade i .

$H(\mathbf{p})$ é a equação que descreve a entropia da informação segundo uma certa distribuição de probabilidade \mathbf{p} (Shannon, 1948). Ela se refere a média de bits necessária, por símbolo, para se representar a informação gerada por um processo estocástico e é expressa como:

$$H(\mathbf{p}) = - \sum_{\alpha} p_{\alpha} \log_2 p_{\alpha},$$

onde p_{α} é a probabilidade estacionária segundo (Page et al., 1999).

5 Assortatividade

A assortatividade é uma medida de quão nós similares estão conectados entre si (Newman, 2003). Quaisquer atributos dos nós podem ser usados como medida de assortatividade, mas a assortatividade de grau é de especial interesse na análise de redes, pois permite caracterizar sua topologia. Redes com assortatividade negativa (desassortativas), remetem à topologia de “eixo e raios” (*hub and spokes*), onde nós de grau muito alto (*hubs*) se conectam a muitos nós de grau mais baixo (Barabási and Pósfai, 2016). A Figura 2a mostra um exemplo de uma rede em que predomina essa topologia.

A assortatividade de grau em uma rede direcionada é caracterizada pelo coeficiente de correlação de Pearson dos graus (Newman, 2003; Barabási and Pósfai, 2016).

6 Análises do MCar Usando Compressão de Fluxo em Grafos

Essa seção apresenta os resultados da aplicação das técnicas descritas nas seções anteriores para identificar padrões relevantes no MCar, especialmente na identificação das Ilhas Ocupacionais e na descrição de suas características.

A distribuição de grau, a conectividade do MCar e sua relação com modelos de carreira são descritos na Seção 6.1. A identificação de Fronteiras de Carreira e sua caracterização são apresentadas na Seção 6.2.

6.1 Distribuição de Graus, Ego-Grafos e Modelos de Carreiras

A rede de ocupações possui a distribuição de grau exibida no gráfico de Pareto da Figura 3a. Ela possui uma grande concentração de conexões em poucos nós, formando *hubs*. Os 25% nós de maior grau são responsáveis por 85,86% das conexões da rede.

Por ser uma rede direcionada, um certo cuidado na análise é necessário, uma vez que o grau dos nós não corresponde diretamente ao número de nós conectados. Imaginando que cada nó pode

ter uma ou duas conexões com o mesmo nó vizinho, o grau k_i , varia em relação ao número de vizinhos entre $k_i = v_i$ (uma conexão por vizinho) e $k_i = 2v_i$ (cada vizinho possui uma conexão de entrada e saída com k), onde v é o número de nós conectados ao nó i por uma conexão de entrada ou saída.

A contribuição de cada nó para a conectividade da rede pode ser observada utilizando-se ego-grafos. Um ego-grafo é um subgrafo formado por um nó, todos os seus vizinhos e as conexões entre eles (Newman, 2010). Como o número de conexões em uma rede direcionada não representa necessariamente o número de nós conectados, o uso do ego-grafo permite identificar *hubs* que possuem maior conectividade sem que necessariamente possuam o maior número de conexões.

No gráfico de Pareto da Figura 3b as ocupações estão ordenadas pelo tamanho do ego-grafo. Exatamente 25% dos maiores ego-grafos são suficientes para conectar 98,31% da rede. O nó de maior grau, *Auxiliar Administrativo*, possui um ego-grafo com 3.369 ocupações das 7.267, ou seja, 46,36% das ocupações de alguma forma possui fluxo de profissionais saindo ou entrando dessa ocupação.

A Tabela 1 contém os 5 maiores ego-grafos do MCar. Devido a presença de *hubs* de grau elevado é esperado que os maiores ego-grafos compartilhem uma grande quantidade de nós.

Ocupação	Ego-grafo	Inéditos
Aux. Administrativo	3.369	3.369
Vendedor	1.790	197
Ass. Administrativo	1.704	89
Recepcionista	1.338	52
Atendente	1.249	32
Auxiliar de Produção	1.061	215

Tabela 1: Tabela com os 5 maiores ego-grafos do MCar. A coluna *Ego-grafo* possui o tamanho do ego-grafo. A coluna *Inéditos* contém o número de nós que o ego-grafo possui que não estão nos ego-grafos anteriores.

Essa característica reforça conceitos sobre movimentação profissional como a Carreira sem Fronteiras e a Carreira Proteana, indicando que uma grande quantidade de profissionais trilha por carreiras variadas e se adaptam a diferentes ocupações. No entanto, há uma rápida queda no número de conexões e no tamanho do ego-grafo, sugerindo que esses conceitos não se aplicam a todas as ocupações igualmente e que um número pequeno delas funciona como “coringa” ou “ponte” para movimentação profissional. Se essa característica está presente por que as ocupações são mais genéricas, por serem típicas do início de carreira ou por algum efeito de mercado é algo a ser investigado em trabalhos futuros.

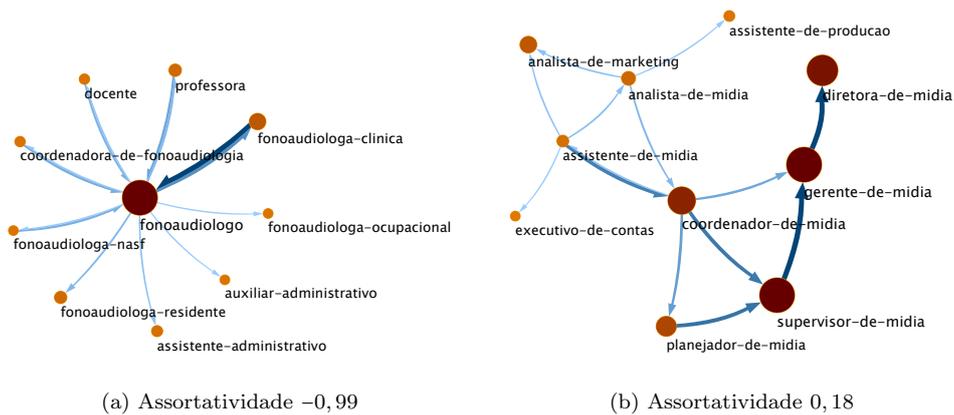


Figura 2: Assortatividade e Topologia

A Figura 3b ilustra a diferença entre o grau e a conectividade da rede que é importante do ponto de vista dessa pesquisa. Enquanto o grau identifica corretamente os *hubs*, o gráfico de Pareto dos ego-grafos mostra que a rede é mais conectada do que seu número de conexões. Comparando os dois gráficos, percebe-se que, enquanto a distribuição de grau e do tamanho do ego-grafo são similares, o acumulado do primeiro cresce mais suavemente que o do segundo, indicando que os *hubs* possuem proporcionalmente mais conexões unidirecionais, ou seja, menos conexões conectam mais nós.

Duas coisas devem ser consideradas aqui: essa característica dos *hubs* pode ser um efeito colateral do corte artificial de 5 conexões; e uma análise similar poderia ser obtida projetando uma rede não-direcionada e analisando sua distribuição de grau.

Observando a Figura 3 é possível notar que cerca de 1/5 das ocupações possui apenas uma conexão (1.547 delas, 21,29%), mas pouco mais de 1/3 (2.682 delas, 36,91%) possui um ego-grafo de tamanho 2. A diferença entre ambos são os ego-grafos que possuem múltiplas conexões, tanto auto-referentes quanto de entrada e saída com seu vizinho.

Isso indica que grande parte das ocupações (1/3) está na “borda” da rede, conectada a ela por apenas um vizinho. Se os *hubs* reforçam os conceitos de Carreiras Proteanas e Carreiras sem Fronteiras, a quantidade de ocupações na borda da rede indica o contrário. A presença de ambas características sugere novamente que as teorias mencionadas não se aplicam a todas as ocupações. Apesar dessa pesquisa conseguir indicar quais ocupações parecem se encaixar melhor nessas teorias e quais não, ela é incapaz de responder quais são as alternativas e por que elas acontecem.

6.2 Identificando Fronteiras de Carreira

Nos experimentos realizados, o algoritmo Infomap foi utilizado para encontrar o particionamento mais

Método	Compressão
Simples	46,08%
Com sobreposição	50,30%

Tabela 2: Métodos de detecção de comunidades e seus resultados. O *Método* se refere ao tipo (variação) de particionamento. *Compressão* significa qual a compressão de informação obtida pelo Infomap em relação a uma rede sem particionamento.

coeso do MCar em relação ao fluxo. Como exposto nas Seções 2 e 4, esse particionamento é a própria definição de fronteiras de carreira.

Para esse algoritmo, quanto maior a compressão da informação do caminho do *random walker*, melhor é o particionamento da rede (Rosvall et al., 2009).

O MCar foi testado em duas variações do algoritmo:

- Simples: consiste na detecção de comunidades em que cada nó pertence a apenas uma única comunidade.
- Com sobreposição: permite-se que nós participem de mais de uma comunidade.

A variação com maior compressão revela a melhor caracterização de particionamento (Viamontes Esquivel and Rosvall, 2011; Rosvall and Bergstrom, 2011; Edler et al., 2017). Se a variação com sobreposição possui maior compressão de informação do que as outras, assume-se que suas comunidades, intrinsecamente, compartilham nós.

Os resultados de compressão para cada variação são apresentados na Tabela 2. É possível observar que há ganho na compressão de informação quando se permite a sobreposição de comunidades, revelando que a movimentação dos profissionais é melhor caracterizada quando as ilhas compartilham ocupações.

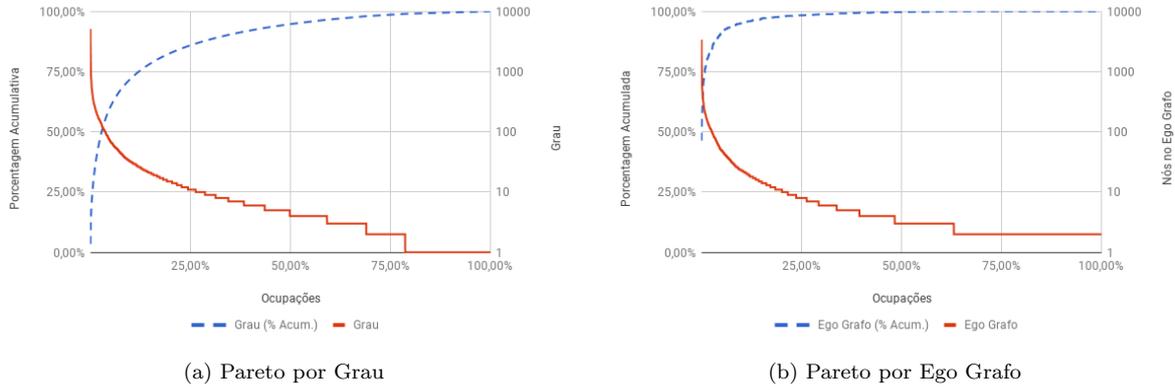


Figura 3: Gráficos de Pareto com a distribuição de grau do MCar e com o tamanho dos ego-grafos. Os gráficos estão colocados lado a lado para facilitar a comparação. Os nós estão ordenados de maneira decrescente no eixo horizontal; na Figura 3a as ocupações de grau maior estão à esquerda e ocupações de menor grau estão à direita. Na Figura 3b elas estão ordenadas pelo tamanho do ego-grafo. Os eixos direitos, em escala logarítmica, estão associados às linhas cheias e representam, respectivamente, o grau do nó e o tamanho do ego-grafo. Os eixos esquerdos estão associados às linhas tracejadas; na Figura 3a ela representa a somatória de todos os graus em porcentagem e qual a contribuição das ocupações à esquerda para essa totalização; na Figura 3b ela representa a quantidade de ocupações que os ego-grafos à esquerda conectam.

Apesar de usar a expressão “compartilham ocupações”, a interpretação mais correta é que as ilhas possuem *suas próprias versões* das ocupações. Uma ocupação que pertence a mais de uma comunidade significa que os profissionais que transitam por ela, em sua maioria, voltam aos seus grupos de origem ao invés de migrarem para outros. Não fosse dessa forma, esses grupos conectados seriam considerados como um conjunto único de ocupações pelo algoritmo ou a ocupação estaria apenas na comunidade com fluxo predominante.

Existem duas explicações possíveis. A primeira é que essas ocupações são *homônimas* e, portanto, distintas, o que significa que há uma fronteira entre elas, como exibido no exemplo da Figura 4. A segunda é que essas ocupações são similares, mas fatores não especificados dificultam a transição de um grupo para outro, o que também indica uma fronteira de carreira, ainda que mais fraca.

Um exemplo da primeira explicação é a ocupação “Coordenador de Projetos”, que está associada tanto à Arquitetura quanto à Software. Um possível exemplo da segunda é a ocupação de “Professor” que aparece em diversas ilhas, sugerindo que há especializações nessa ocupação que impedem que um Professor se movimente entre ilhas, o que condiz com o conhecimento tácito sobre a profissão, um Professor de Matemática não se transforma em um Professor de Língua Portuguesa facilmente.

Em qualquer uma dessas interpretações, as ocupações que participam de mais de uma ilha ocupacional revelam uma fronteira e não uma ponte.

As comunidades, em sua maioria, possuem uma topologia em que predomina a forma de estrela ou estrelas com múltiplos centros. Isso significa

poucos *hubs* conectando quase todos os outros nós. Esse formato é caracterizado por uma assortatividade negativa, ou seja, nós de alto grau conectados a nós de baixo grau (Barabási and Pósfai, 2016).

A assortatividade das 624 ilhas com ao menos três ocupações é exibida na Figura 6. Quase todas elas possuem assortatividade negativa, reforçando o conceito de polos ocupacionais. Na Figura 7 são apresentados três exemplos que ilustram a importância dos polos ocupacionais.

A ilha com o polo “Aluno de Iniciação Científica” é exibida na Figura 7a. A topologia estrelada é claramente observável, irradiando-se a partir do *polo ocupacional*. O fluxo predominante parte de *Aluno de Iniciação Científica*, passa por *Aluno de Mestrado*, *Aluno de Doutorado* e termina em *Pos-Doutorando* sem que haja fluxo na direção contrária, reforçando o conhecimento tácito sobre a área.

O polo ocupacional é claramente *Aluno de Iniciação Científica* que recebe um forte fluxo de *Monitor*. O reforço do polo deve distribuir o fluxo de pessoas para as outras ocupações. Outro ponto de especial interesse é que, a partir de *Aluno de Mestrado* o fluxo intensifica até *Pos-Doutorando*, sugerindo que os profissionais que se tornam *Alunos de Mestrado* têm forte tendência de prosseguir na carreira acadêmica.

Por outro lado, o enfraquecimento do polo tende a minar o fluxo de pessoas para outras ocupações, provocando o enfraquecimento da ilha como um todo.

Essa análise pontual dá dimensão à contribuição desse trabalho. Nesse caso, identificada a ilha, o polo e o fluxo predominante, ações para reforçar

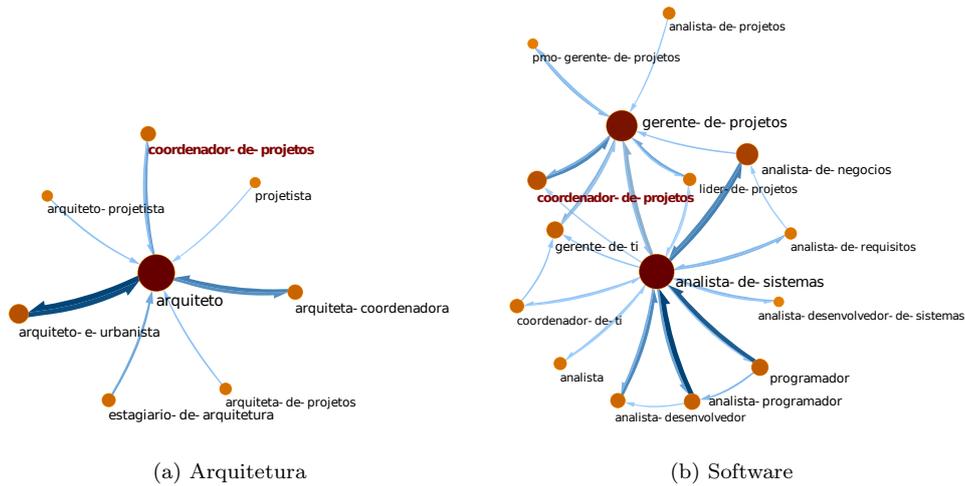


Figura 4: Duas ilhas ocupacionais compartilhando a ocupação “Coordenador de Projetos” (em destaque). Apesar de possuírem o mesmo nome são ocupações diferentes, a movimentação entre as ilhas é pouco representativa se comparada a movimentação intra-ilha. Apenas uma fração das ocupações de cada ilha está exibida nessa figura. O tamanho do nó indica o fluxo de profissionais que passa pela ocupação enquanto a espessura da conexão indica o volume do fluxo entre elas.

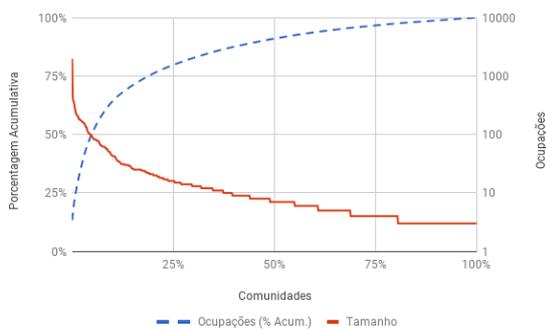


Figura 5: Gráfico de Pareto com a distribuição de tamanho das ilhas ocupacionais. As comunidades estão ordenadas pelo número de ocupações que contém, ilhas maiores estão à esquerda e menores à direita. O eixo direito, em escala logarítmica, está associado à linha cheia e representa o número de ocupações da ilha. O eixo esquerdo está associado à linha tracejada e representa o total de ocupações e qual a contribuição acumulativa das ilhas à esquerda do ponto para ela.

uma ocupação, como *Pos-Doutorando*, por exemplo, passam a ser óbvias. Essa análise superficial sugere que aumentar o número de alunos de iniciação científica, desestimular sua movimentação para outras ocupações ou incentivar o fluxo para o mestrado provocaria o efeito desejado. A proporção de pessoas em cada fluxo fornece um “funil” para se estimar quantas pessoas seriam necessárias em cada ocupação para atingir um certo resultado.

A ilha ocupacional relacionada à *Recursos Humanos*, exibida na Figura 7b, possui uma topologia de estrela com polo em *Analista de Recursos Huma-*

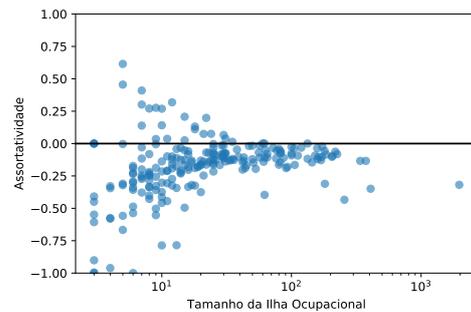


Figura 6: Assortatividade pelo número de ocupações na ilha com o eixo x em escala logarítmica.

nos que se alimenta principalmente de *Assistente de Recursos Humanos* e *Analista de Departamento Pessoal*. Apesar de alguma dispersão, o fluxo prossegue para ocupações mais administrativas *Coordenador*, *Supervisor* e *Gerente*, com uma certa força para *Consultor*.

Nessa ilha, o polo sugere ser a própria razão de ser da área, uma vez que as ocupações anteriores não estão diretamente relacionadas à função ou são juniores e o fluxo prossegue para ocupações administrativas. As outras ocupações *Analistas* que se dispersam a partir do polo sugerem especializações do *Analista de Recursos Humanos*, em especial *Analista de Administração de Pessoal* e *Analista de Recrutamento e Seleção*, que possuem fluxos similares de entrada e saída para ele. Peculiarmente, *Psicólogo* possui fluxo similar, mas as funções esperadas da ocupação não são diretamente relacionadas à *Recursos Humanos*.

No último exemplo na Figura 7c gira em torno do polo *Analista Ambiental*, mas não possui o

mesmo fluxo claro de progressão como os exemplos anteriores. Junto a ele, outros dois polos menores *Engenheiro Ambiental* a *Técnico em Meio Ambiente* trocam fluxos com o polo, enquanto ocupações administrativas como *Coordenador*, *Gerente* e *Consultor* recebem fluxos dos três polos enquanto *Estagiário* as alimenta.

Ao lado esquerdo da Figura 7c é possível observar ocupações de ensino e pesquisa, como *Professora*, *Docente* e *Pesquisador*. Sua presença na ilha sugere que são especializadas na área Ambiental, ainda que os profissionais não tenham deixado explícito em seus currículos.

A presença de ocupações que se relacionam com *Segurança do Trabalho* e *SMS* (Segurança, Meio Ambiente e Saúde) indicam um certo compartilhamento dessa área e Meio Ambiente, ainda que não seja forte o suficiente para juntar as ilhas. São ocupações que, apesar de compartilhar o título, parecem ser executadas em contextos diferentes, de modo que a ocupação *Engenheiro de Segurança do Trabalho* possui pessoas que vão e vêm da área ambiental, sem haver uma troca com outras áreas.

Essa característica condiz com o conhecimento tácito de que é mais fácil permanecer em uma certa área de atuação (Meio Ambiente), ainda que a profissão em si seja mais geral (Segurança do Trabalho).

Os exemplos acima ilustram como essa pesquisa pode ser utilizada. Ela contribui fortemente para as análises feitas, mas precisam de conhecimento externo, como o significado das funções (“Gerente é uma ocupação administrativa”), para dar significado aos resultados.

A pesquisa consegue delinear características essenciais da movimentação profissional, como ilhas e polos, mas é insuficiente, por si, para explicar o porque elas ocorrem e o que exatamente significam em cada contexto.

7 Conclusões e Perspectivas Futuras

Esse trabalho reúne os conceitos fronteiras de carreira de (Gunz et al., 2007), os trabalhos de Ciência de Redes sobre detecção de comunidades por fluxo baseados em (Rosvall et al., 2009) e os dados de um dos maiores sites brasileiros de carreira (VAGAS Tecnologia, 2014) para contribuir na compreensão da movimentação profissional.

A rede profissional possui *hubs*, o que fornece indícios de que carreiras menos regulares são comuns, como advogam as Carreiras sem Fronteiras e Carreiras Proteanas. Ao mesmo tempo a presença de muitos pequenos ego-grafos e ilhas ocupacionais diminutas sugerem que há alternativas para essas teorias.

A Equação de Mapa (Rosvall et al., 2009) foi aplicada usando o software Infomap (Edler and Rosvall, 2012) para identificar as fronteiras de carreiras, gerando *ilhas ocupacionais*, que são conjun-

tos de ocupações em que a movimentação interna é significativamente mais frequente do que movimentações para outras *ilhas*. O termo foi introduzido para facilitar a discussão e como uma analogia ao conceito sugerido por (Abbott, 1995) em que a fronteira define o grupo, ao invés do grupo definir a fronteira. Essa abordagem usa a diferença entre os grupos para defini-los ao invés de alguma característica intrínseca.

A análise da assortatividade das ilhas mostrou que sua topologia tende ao formato estrelado ou de eixo e raios, o que motivou a introdução do termo *polos ocupacionais* para identificar ocupações que dão coesão à ilha. Alguns exemplos emblemáticos dão dimensão aos dois conceitos introduzidos nesse trabalho.

Conclui-se na expectativa que esse trabalho contribua para a compreensão das transições profissionais, no entanto, é preciso reconhecer as limitações dessa proposta e as questões que permanecem abertas.

Os aspectos sociais e psicológicos das fronteiras de carreira não foram abordados e são eles quem podem responder a perguntas como “por que as fronteiras estão onde estão?” e “por que as ilhas possuem essa topologia?”. Espera-se que o conteúdo apresentado contribua de maneira quantitativa em discussões sociais e psicológicas sobre movimentação profissional.

Outros trabalhos podem aprofundar a questão da dinâmica da rede, focando em modelos que possam prever os efeitos esperados quando uma ocupação ganha ou perde atratividade, ou quando um fluxo é incentivado ou estrangulado. Esse trabalho ensaia um começo tímido nessa direção ao propor o conceito de polos ocupacionais e identificar sua importância nas ilhas de ocupações.

Outra linha de pesquisa atrativa está em comparar o MCar com redes aleatórias geradas a partir de características similares à da rede real, procurando identificar quais delas resultam em observações comparáveis. Essas características indicam possíveis explicações para os comportamentos encontrados a partir de uma abordagem puramente estatística (Barabási and Pósfai, 2016).

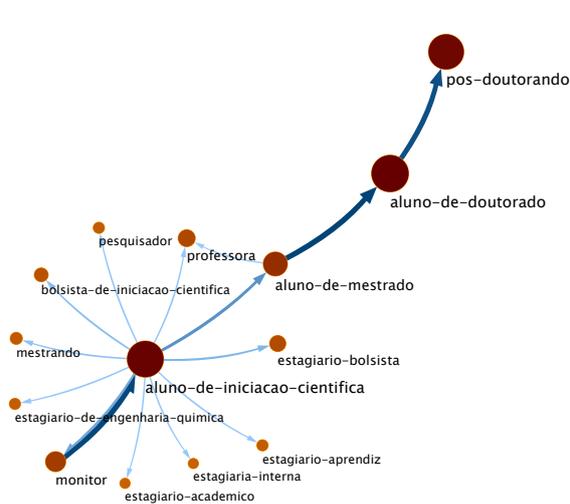
O MCar é uma fonte considerável de informação sobre o mercado de trabalho brasileiro. Outras pesquisas sobre profissões e carreira, em especial utilizando técnicas de Ciência de Redes, como *motifs*, podem contribuir para uma melhor compreensão da movimentação profissional.

8 Agradecimentos

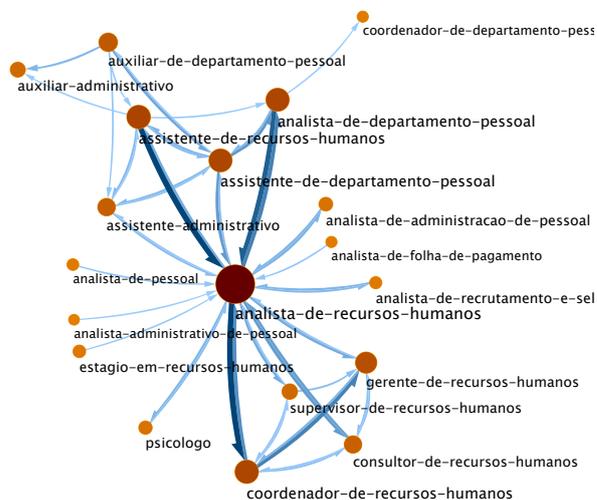
Os autores agradecem ao CNPq, Capes, Fapesp e MackPesquisa pelo auxílio financeiro. O Laboratório de Computação Natural e Aprendizagem de Máquina agradece à Intel pelo reconhecimento como um Centro de Excelência em Inteligência Artificial.

Referências

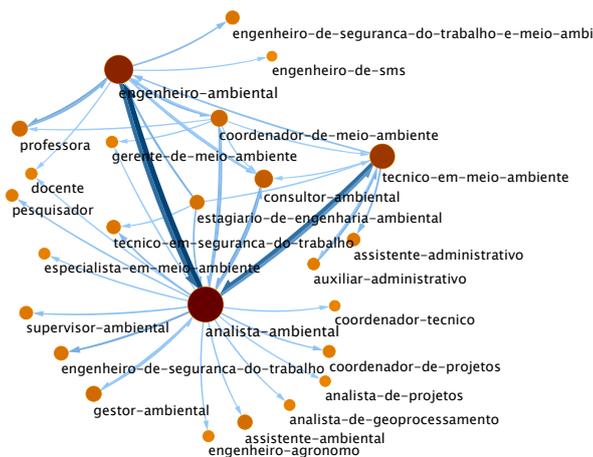
- Abbott, A. (1995). Things of boundaries, *Social Research* **62**(4): 857–882.
- Arthur, M. B., Hall, D. T. and Lawrence, B. S. (1989). Generating new directions in career theory: the case for a transdisciplinary approach, in M. B. Arthur, D. T. Hall and B. S. Lawrence (eds), *Handbook of Career Theory*, Cambridge University Press, pp. 7–25.
- Barabási, A.-L. and Pósfai, M. (2016). *Network Science*, 1 edn, Cambridge University Press.
- Baruch, Y. (2004). Transforming careers: from linear to multidirectional career paths: Organizational and individual perspectives, *Career Development International* **9**(1): 58–73.
- Bendassolli, P. F. (2009). Recomposição da relação sujeito-trabalho nos modelos emergentes de carreira, *RAE-Revista de Administração de Empresas* **49**(4).
- Edler, D., Bohlin, L., Rosvall and Martin (2017). Mapping Higher-Order network flows in memory and multilayer networks with infomap, *Algorithms* **10**(4): 112.
- Edler, D. and Rosvall, M. (2012). The MapEquation software package, <http://www.mapequation.org>. Acessado em 2017-6-10.
- Grünwald, P. D. (2007). *The Minimum Description Length Principle*, MIT Press.
- Gunz, H. P., Peiperl, M. A. and Tzabbar, D. (2007). Boundaries in the study of career, in H. P. Gunz and M. A. Peiperl (eds), *Handbook of Career Studies*, 1 edn, SAGE Publications, Inc, pp. 471–494.
- Morgado, F. (2017). *Silvio Santos*, 1 edn, Matrix.
- Newman, M. (2003). Mixing patterns in networks, *Physical Review E*.
- Newman, M. (2010). *Networks: An Introduction*, 1 edn, Oxford University Press.
- Page, L., Brin, S., Motwani, R. and Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web.
- Rosvall, M., Axelsson, D. and Bergstrom, C. T. (2009). The map equation, *Eur. Phys. J. Spec. Top.* **178**(1): 13–23.
- Rosvall, M. and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems, *PLoS One* **6**(4): e18209.
- Shannon, C. E. (1948). The bell system technical journal, *A mathematical theory of communication* **27**: 379–423.
- Steiner, N. C. (2009). *Monja Coen*, 1 edn, Mescla.
- Sullivan, S. E. and Baruch, Y. (2009). Advances in career theory and research: A critical review and agenda for future exploration, *J. Manage.* **35**(6): 1542–1571.
- VAGAS Tecnologia (2014). Mapa VAGAS de carreiras, <http://www.vagas.com.br/mapa-de-carreiras/>. Acessado em 2017-9-23.
- Viamontes Esquivel, A. and Rosvall, M. (2011). Compression of flow can reveal Overlapping-Module organization in networks, *Phys. Rev. X* **1**(2): 021025.



(a) Iniciação Científica



(b) Recursos Humanos



(c) Ambiental

Figura 7: Três exemplos de ilhas ocupacionais. O tamanho e a cor do nó estão relacionados ao fluxo de profissionais que passa por ele, quanto maior e mais escuro o nó, maior o fluxo. Da mesma forma, a espessura e a cor das conexões está relacionada ao fluxo de profissionais entre uma ocupação e outra. O polos ocupacionais são identificados pelo volume de pessoas transitando por eles e pela sua alta conectividade (visualmente os nós maiores e mais centrais). Para facilitar a visualização, nas Figuras 7b e 7c, apenas as conexões e nós de maior fluxo são apresentados. A Figura 7a, no entanto, mostra a ilha completa.