

### 3 O MUNDO DAS REDES

“O átomo é o passado. O símbolo da ciência para o próximo século é a rede dinâmica” (KELLY, 1994, p.25, tradução nossa). “Quando no futuro se fizer o elenco das abstrações que marcaram mais profundamente o espírito humano no decorrer desta segunda metade do século [XX], ter-se-á, certamente, a «cifra lógica» [...], mas, também, quase tão importante como aquela, ter-se-á o objecto «rede»” (ROSENSTIEHL, 1988, p.229).

Exageradas ou não, estas previsões dialogam com o contexto exposto no capítulo anterior e com a explosão de publicações verificada a partir da última década do século XX sobre a teoria de redes e suas aplicações.

Nesse contexto, importantes semelhanças topológicas têm sido reveladas em redes de natureza tão diferentes como as conexões neuronais, as amizades entre usuários do *facebook*, as interligações entre dispositivos de um circuito elétrico, os hipertextos da *web* e os mapeamentos de atuações conjuntas em filmes de cinema ou de coautorias de artigos científicos.

Assim também, textos têm sido investigados enquanto redes de palavras e expressões que, em suas frequências de uso ou articulações, revelam características linguísticas, culturais e autorais.

Como em qualquer campo específico de conhecimento, há toda uma linguagem própria associada aos estudos dessas redes, sendo exemplo disso termos como *redes de pequeno mundo*, *redes livres de escala*, *centralidades de grau*, *de intermediação* e *de proximidade*, *hubs*, *clusters*, *mineração* e *lematização de textos*.

É com o objetivo de descrever esses e outros conceitos, bem como apresentar uma visão panorâmica das pesquisas sobre as chamadas redes complexas que se apresenta este capítulo. Diferentemente de uma extensa revisão bibliográfica, optou-se por destacar alguns exemplos que ao mesmo tempo em que pudessem ser articulados entre si, conformando a segunda grande rede abordada neste trabalho, também fossem capazes de estabelecer conexões com os demais capítulos desta tese.

### 3.1 NÓS, LINKS E SUAS PROPRIEDADES

Grosso modo uma rede é um conjunto de objetos (vértices ou nós) conectados (por arestas ou *links*) de alguma maneira. A partir dessa generalização, uma série de definições específicas pode ser utilizada para analisar e distinguir os mais diversos tipos de redes. A mais básica delas talvez seja o grau de um nó, definido pelo número de *links* a ele conectado. No *facebook*, por exemplo, o grau de cada pessoa corresponde ao número de amigos que ela possui. Algumas pessoas se destacam, tendo um número de amigos em escala muito superior ao da maioria dos usuários; são os *hubs* da rede.

Por vezes o que mais chama a atenção não é o número de amigos que uma pessoa tenha, mas a alta conectividade entre seus amigos, caracterizando uma região de alta aglomeração. Se todos os meus amigos forem amigos uns dos outros, diz-se que meu coeficiente de aglomeração ou *cluster* vale 1. Na situação oposta, quando meus amigos não se conhecem em absoluto, meu *cluster* vale 0.

Medida de grande importância é a distância entre dois nós de uma rede. Por distância, entenda-se um conceito topológico e não algo que se possa medir com fita métrica. Assim, no *facebook* a distância entre dois usuários é definida como o caminho mais curto entre eles, ou seja, o menor número de *links* que se deve percorrer para sair de um e chegar ao outro. Desse modo a distância entre dois amigos vale 1 e a minha distância para o amigo de meu amigo vale 2. O caminho mais curto, ou *caminho geodésico*, é também utilizado na definição do diâmetro de uma rede que corresponde ao maior *caminho geodésico* nela existente.

Quanto maior a rede, maior a importância de se realizem medidas comparativas entre seus nós. Claro que a forma mais elementar dessa comparação é por meio dos cálculos dos valores médios de graus, *clusters* e *caminhos geodésicos*. No entanto, da mesma forma que a simples média das notas de uma turma de alunos não nos traz informação lá muito relevante, no estudo de grandes redes, é padrão averiguar a distribuição daquelas medidas ao longo de toda a rede.

A distribuição de grau, por exemplo, é definida como uma fração do total de nós,  $n$ , da rede que têm determinado grau,  $k$ :  $P(k) = n_k/n$ . Se em uma festa com 100 pessoas ( $n = 100$ ), verifica-se que 5 ( $n_k = 5$ ) delas têm 10 amigos ( $k = 10$ ), ao passo que outras 10 ( $n_k = 10$ ) tem apenas 5 amigos ( $k = 5$ ), então as distribuições de grau 10 e 5 valem, respectivamente,  $P(10) = 0,05$  e  $P(5) = 0,1$ . A tabela 1 apresenta mais alguns valores correspondentes a este exemplo.

Tabela 1 – Eventual distribuição estatística do número de pessoas com  $k$  amigos em uma festa com 100 participantes ( $n = 100$ ).

Nº de amigos/grau ( $k$ )	Nº de pessoas com $k$ amigos ( $n_k$ )	Distribuição $P(k) = n_k/n$
10	5	0,05
5	10	0,1
3	15	0,15
2	20	0,2
1	50	0,5
<b>Total</b>	<b>100</b>	<b>1</b>

FONTE: Próprio autor

A realização de cálculos como esses dão hoje suporte para explicações de algumas propriedades comuns a muitas redes, tal qual o popular fenômeno do mundo pequeno ou dos seis graus de separação. Em fins da década de 60, Milgram (1967) fez um experimento que ganhou notoriedade a esse respeito. Se alguém quisesse enviar um recado para uma pessoa desconhecida escolhida ao acaso, quantos contatos deveriam ser feitos até que esse recado, passado de pessoa a pessoa, chegasse ao alvo? Contextualizada em uma região dos Estados Unidos, a pesquisa de Milgram mostrou que bastaria um máximo de seis contatos.

Sem qualquer proeminência mágica do número seis, de lá para cá, diversas pesquisas reconheceram que as redes sociais são, de fato, caracterizadas por valores relativamente baixos para a distância média entre seus nós. Fosse essa a única característica relevante, se poderia imaginar que as amizades formassem laços aleatórios ao longo da rede, encurtando a distância média entre as pessoas. Mas aferições das relações entre os amigos de cada pessoa pertencente à rede revelam alto valor para seu coeficiente de aglomeração médio.

A combinação dessas duas propriedades, alta aglomeração e pequena distância média entre os nós, é o que se convencionou como fatores de identificação das *redes de pequeno mundo*, ou *redes de Watts e Strogatz*, em referência aos pesquisadores que anunciaram a descoberta em artigo publicado na Nature em 1998 (WATTS; STROGATZ,1998). Tal revelação, aliás, já se fez extrapolando os limites das redes sociais, com a apresentação naquele artigo dos dados expostos na tabela 2 relativos a três redes de diferentes naturezas: atores de cinema ligados por sua atuação conjunta; subestações, geradores e transformadores conectados por linhas de transmissão de alta voltagem; neurônios do verme *Caenorhabditis elegans* (*C. elegans*) com *links* estabelecidos por sinapses ou *junções gaps*.

Tabela 2 – As três redes de “pequeno mundo” identificadas por Watts e Strogatz em trabalho pioneiro publicado em 1998.

Rede	Nº de nós	Grau médio	Distância média entre nós	Aglomerção média ( <i>cluster</i> )
Atores de cinema (1)	225.226	61	3,65	0,79
Rede elétrica (2)	4.941	2,67	18,7	0,08
<i>C. elegans</i>	282	14	2,65	0,28

FONTE: WATTS; STROGATZ, 1998.

(1) Dados armazenados na *Internet Movie Database* (IMDB) em abril de 1997;

(2) Dados disponibilizados em 1996 por departamento de engenharia elétrica do oeste norte-americano.

É importante salientar que caracterizar uma rede como de pequeno mundo, não implica em necessária proximidade absoluta entre seus componentes. Quando comparada à rede de atores de cinema, a rede elétrica exposta na tabela 2 apresenta grande distância média entre seus nós e baixo coeficiente médio de aglomeração. Trata-se, todavia, de uma rede com nós de poucos *links*, naturalmente menos densa. A comparação correta a ser feita é com o modelo de redes construídas aleatoriamente que prevê coeficientes de aglomeração bem mais baixos.<sup>4</sup> Além disso, identificar que seja pequena a distância média entre os nós de uma rede, não é indicativo de facilidade quanto à transmissão de informações entre dois de seus nós. Voltando ao *facebook*, estar a dois graus de separação de outro usuário somente resultará na transmissão de uma mensagem por esse caminho mais curto se eu souber de antemão que ele é amigo de meu amigo. Caso contrário, há milhares de outros caminhos possíveis pelos quais a mensagem pode ficar navegando na rede, sendo o gerúndio aqui proposital.

Outro exemplo, agora relativo à rede de atores de Hollywood. Um ator novato que faça uma ponta em um filme estrelado por Kevin Bacon certamente estará a dois graus de separação de Tom Hanks pela atuação conjunta destas estrelas em *Apollo 13*. Mas isso para nada quer dizer que ele terá grande chance de atuar alguma vez com Hanks. A referência a Kevin Bacon não é fortuita e, sim, motivada pela existência do *Oráculo de Bacon*<sup>5</sup>, página eletrônica que fornece o *número de Bacon* de cada ator presente na base de dados do Internet Movie Database (IMDB). O *número de Bacon* de Penélope Cruz é, por exemplo, 2, haja vista

<sup>4</sup> *Redes aleatórias*, ou *redes de Erdős e Rényi*, são construídas mediante algoritmo em que todos os nós têm a mesma probabilidade de serem conectados entre si. Como consequência espera-se que as conexões sejam distribuídas “democraticamente” (*distribuição de Poisson*), com a prevalência de baixos valores de aglomeração. Caso as três redes citadas se configurassem como aleatórias, seus coeficientes de aglomeração teriam os seguintes valores: 0,00027 para a rede de atores de cinema, 0,005 para a rede elétrica e 0,05 para os neurônios da *C. elegans*.

<sup>5</sup> Funcionalidade disponível em <http://oracleofbacon.org/> (Acesso em 31 Mar. 2014)

que ela atuou em *Sahara* com Willian H. Macy que participou de *Murder in the First* com Kevin Bacon.

A American Mathematical Society<sup>6</sup> e a Microsoft<sup>7</sup> disponibilizam funcionalidades semelhantes para se descobrir o *número de Erdöz*, isto é, o grau de separação para com publicações realizadas pelo destacado matemático Paul Erdöz, falecido em 1996. Na figura 10 se identifica como 5 o *número de Erdöz* de César Lattes.

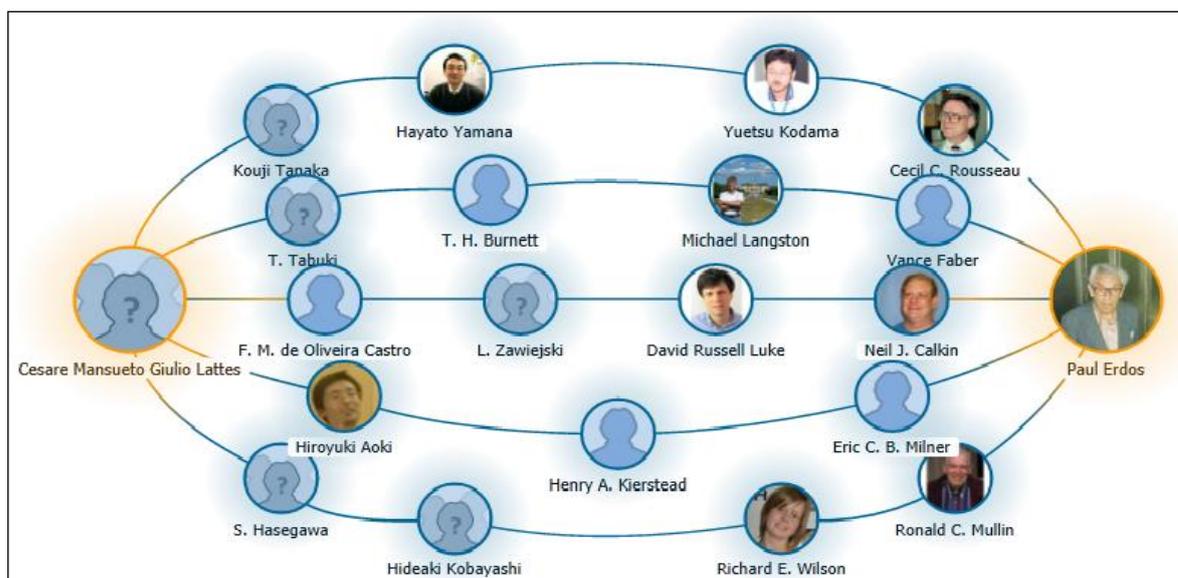


Figura 10 - *Número de Erdöz* de César Lattes

Pouco tempo depois de publicado o trabalho de Watts e Strogatz, seria a vez de Barabási e Albert (BARABÁSI; ALBERT, 1999) anunciarem a percepção de nova propriedade presente em diversas redes: enquanto alguns poucos nós possuem muitos *links*, caracterizando-se como *hubs*, a grande maioria tem poucas conexões. Mais que isso, a distribuição de graus parecia seguir uma lei de potência, com Barabási e Albert cunhando o termo *redes livres de escala* e propondo um modelo matemático para sua formação com base ao processo indicado na figura 11 (BARABÁSI, 2002, p.87), em que nós com mais conexões têm maior probabilidade de cooptarem novos nós acrescidos à rede, algo como a regra de que “os ricos ficam mais ricos”.

<sup>6</sup> Funcionalidade disponível em <<http://www.ams.org/mathscinet/freeTools.html>> (Acesso em 31 Mar. 2014)

<sup>7</sup> Funcionalidade disponível em <<http://academic.research.microsoft.com/VisualExplorer/#2019273&1112639>> (Acesso em 31 Mar. 2014)

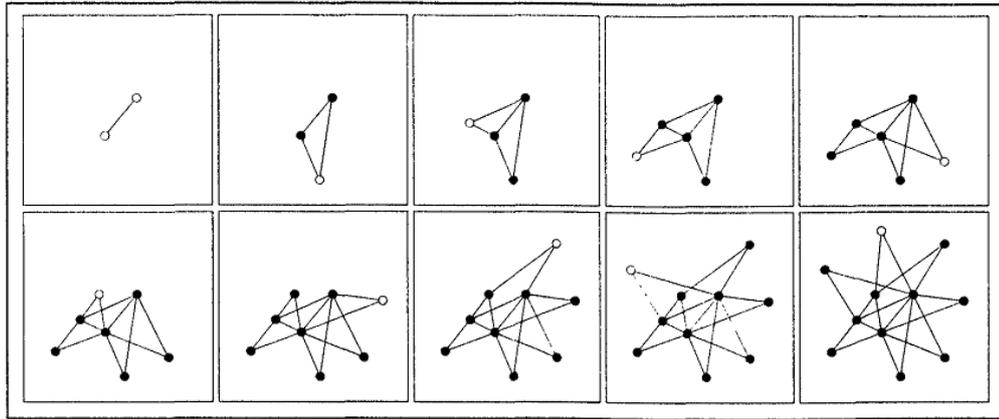


Figura 11 - Formação de uma rede sem escala: a partir da conexão entre dois nós inaugurais (quadro superior esquerdo), cada novo nó (círculo vazio) acrescido à rede tem favorecida sua conexão com os nós que já possuem mais *links*, resultado na formação de *hubs*.

Colocando o algoritmo em ação, este modelo resulta em uma distribuição de graus regida por uma lei de potência,  $N(k) \sim k^{-\gamma}$ , com  $\gamma$  determinando a dimensão da desigualdade presente na rede. Em uma rede com  $\gamma = 2$ , por exemplo, se 400 nós tiverem um único *link*, existirão outros 100 nós com 2 *links*, 25 nós com 4 *links*, 16 nós com 5 *links*, 4 nós com 10 *links* e um único nó com 20 *links*. Como em qualquer fenômeno regido por leis de potência não há, portanto, um valor médio esperado para o grau de um nó, sendo esta a motivação para a atribuição do título de *redes livres de escala*. A figura 12 representa graficamente tal desigualdade nas escalas decimal e logarítmica.

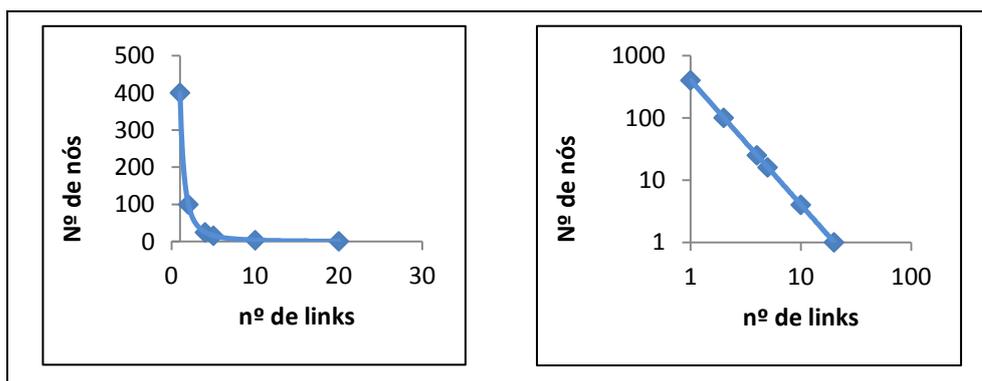


Figura 12 - Em uma *rede livre de escala* com  $\gamma = 2$  ( $N(k) \sim k^{-2}$ ), se 400 nós tiverem apenas um “amigo”, é porque existe um nó que concentra 20 “amizades”. A gradual queda do número de nós conforme aumento do número de *links* é representada por uma curva na escala decimal (à esquerda) e por uma reta na escala logarítmica (à direita). Vale observar que o coeficiente angular desta reta é dado pelo valor do expoente que caracteriza a lei de potência ( $-\gamma$ ), sendo prática comum o uso desta escala nos estudos comparativos entre redes reais e o modelo da *rede livre de escala*.

Apesar de sua popularidade acadêmica, refletida na publicação de inúmeras pesquisas voltadas à medição do expoente  $\gamma$  das mais diversas redes, o modelo das redes livres de escala tem sofrido aperfeiçoamentos e contestações. Há redes, como a *internet*, em que alguns nós mais antigos desaparecem (páginas eletrônicas tiradas do ar pelo próprio autor), ao passo que outros mais jovens ganham repentinamente destaque de *super hubs* (como o Google), fazendo-se necessário ajustar parâmetros ao modelo. Mas também há redes em que a desigualdade entre *hubs* e demais nós não seguem leis de potência.

Além disso, na análise das características de uma rede não basta reconhecer como se distribuem os graus de seus nós, mas também qual a relativa importância de cada um deles. Tal qual ilustrado na figura 13, a depender das conexões entre os nós, pode ocorrer que um nó com poucos *links* seja mais decisivo para a coesão da rede do que um grande *hub*.

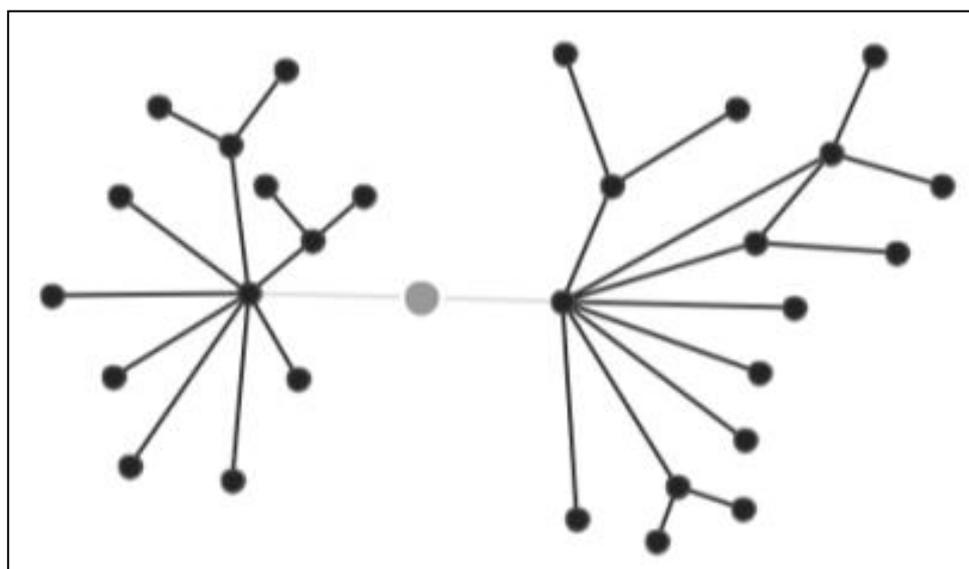


Figura 13 - Apesar de ter apenas dois *links*, o nó cinza claro apresenta vital importância para a coesão da rede.

São várias as grandezas, geralmente chamadas de centralidades, encontradas na literatura da teoria de redes, tais como (a) a *centralidade de grau*, vinculada ao número de *links* de um nó, (b) a *centralidade de intermediação*, determinada pelo número de vezes que um nó se apresenta ao longo do caminho mais curto entre dois outros nós e (c) a *centralidade de proximidade*, associada à distância total de um nó a todos os demais nós. Voltando ao *facebook*, quanto mais amigos uma pessoa tiver, quanto mais sua página for realizada para transmitir informações entre dois desconhecidos e quanto menos intermediários ela precisar para se conectar ao conjunto dos usuários da rede, serão respectivamente maiores suas centralidades de grau, intermediação e proximidade. A título de exemplo, a figura 14

apresenta uma mesma rede sob a perspectiva de cada uma dessas três centralidades, com os nós C e D se destacando como os únicos a possuírem simultaneamente os maiores valores dessas três medidas.

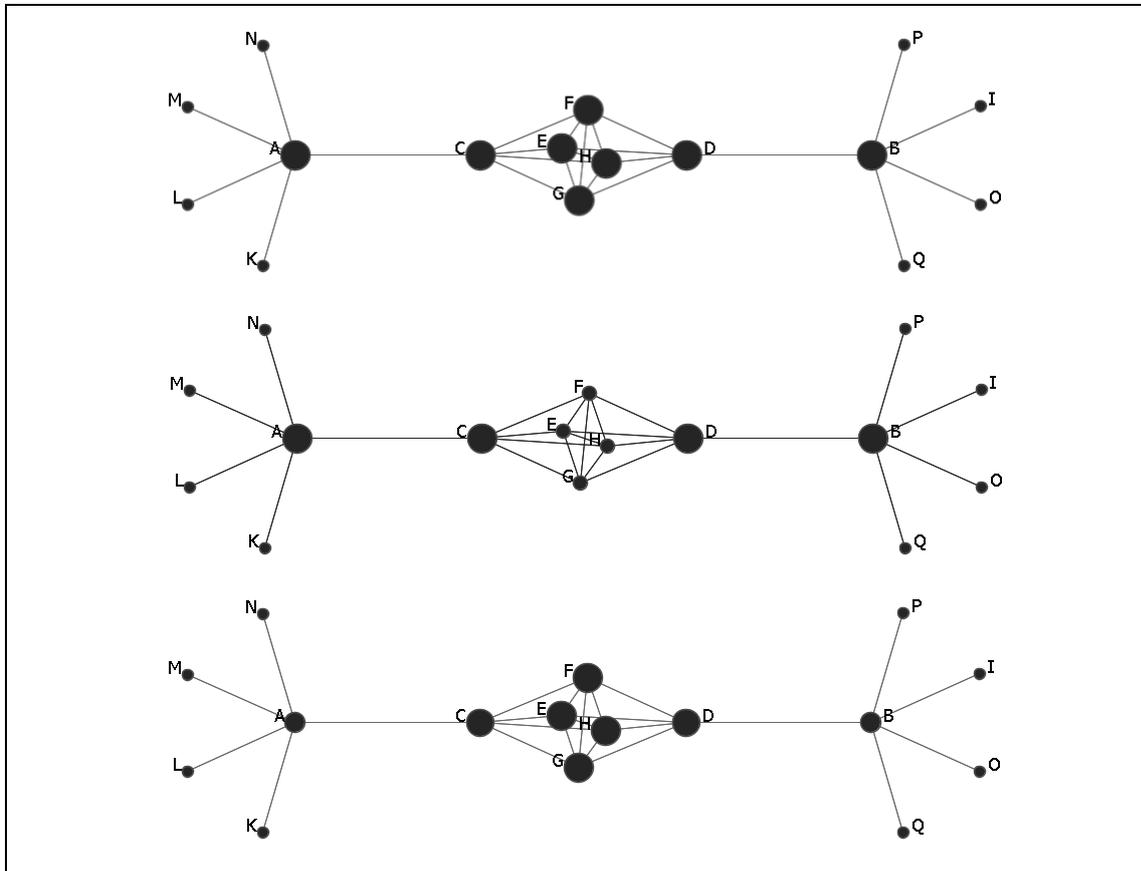


Figura 14 - Uma mesma rede representada com o diâmetro dos nós proporcionais aos valores das centralidades de grau (figura superior), intermediação (figura central) e proximidade (figura inferior). Apesar de terem os mesmos 5 links que os nós E, F, G e H, os nós A, B, C e D possuem maior centralidade de intermediação, com C e D também se apresentando entre os nós de maior centralidade de proximidade.

A importância relativa de cada nó depende, portanto, de sua capacidade para contribuir à eficácia da rede, com nós que absorvem e processam mais informações relevantes podendo ser considerados como centros. Tais centros, no entanto, em nada se assemelham com o nó central de uma teia de aranha. Não só não há um único centro, como também os centros não são fixos, tratando-se de uma estrutura dinâmica em permanente processo de reestruturação com criação e eliminação de nós, transformação de nós em centros, de centros em nós e redes internas que competem ou cooperam entre si. O todo é mais do que a soma das partes. É por investigar sistemas com tais características que o campo de estudo aqui panoramicamente apresentado é comumente chamado de *Teoria das Redes Complexas*.

Comento, por fim, algumas representações visuais comumente utilizadas para as conexões entre os nós de uma rede. A figura 14 mostra três representações, dentre diversas possíveis, de uma mesma rede constituída por 15 nós e 34 *links*, com a primeira delas (figura 15A) posicionando os nós de forma hierárquica a partir daquele de maior grau, a segunda (figura 15B) de acordo com a simetria circular e a terceira (figura 15C) utilizando como modelo um sistema de massas (nós) e molas (*links*) configurado no estado de menor energia possível.

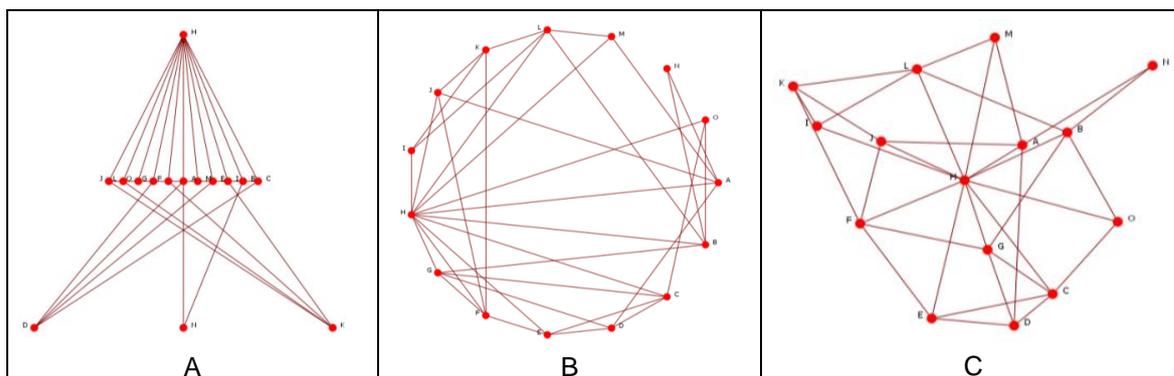


Figura 15 – Diferentes representações de uma mesma rede com 15 nós e 34 *links*, com (A) nós posicionados hierarquicamente a partir do nó H de maior grau, (B) distribuídos por simetria circular e (C) com uso de algoritmo baseado na modelagem de massas (nós) e molas (*links*) configuradas no estado de menor energia possível.

### 3.2 REDES BIOLÓGICAS

A interação entre proteínas que compõem um complexo proteico é um dos vários sistemas biológicos que têm sido investigados no contexto da teoria de redes. Com a teia construída a partir das proteínas como nós e suas interações como *links*, tais redes têm sido mapeadas em organismos tal qual a levedura *Saccharomyces cerevisiae* (*S. cerevisiae*)<sup>8</sup>, resultando em representações como a exposta na figura 16. Análises destes mapeamentos têm apontado para a existência de uma distribuição desigual dos graus de cada nó, com algumas proteínas cumprindo a função de *hubs* e tendo maior probabilidade de serem essenciais ao organismo e de levá-lo à morte em caso de suas exclusões. Segundo Barabási et al. (2001), tal percepção infere a importância de propriedades topológicas na reconhecida robustez contra mutações apresentada pelas leveduras:

<sup>8</sup> Fungo unicelular comumente utilizado como fermento biológico na produção do pão e da cerveja.

A correlação entre a conectividade e a indispensabilidade de uma determinada proteína confirma que, apesar da importância da função bioquímica individual e redundância genética, a robustez contra mutações em levedura também é derivada da organização das interações e posições topológicas das proteínas. (BARABÁSI et al., 2001, tradução nossa).

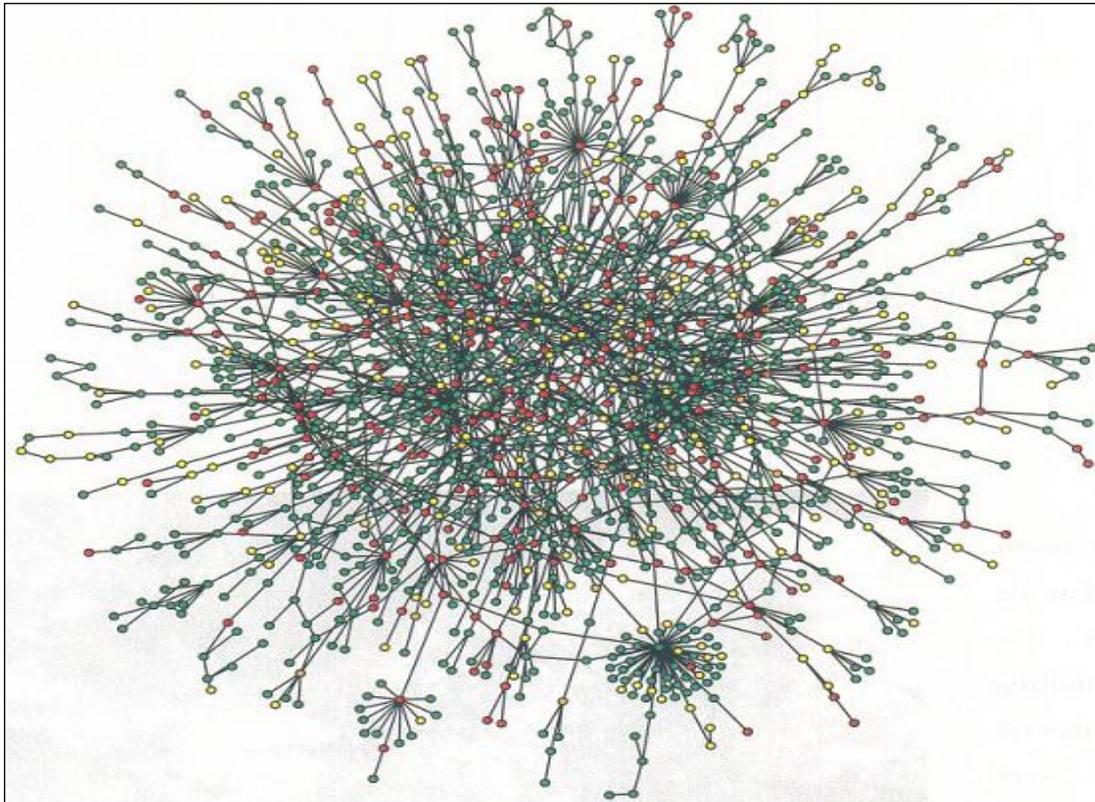


Figura 16 - Mapeamento das interações entre proteínas na levedura *S. cerevisiae* com coloração indicativa da importância para com a sobrevivência do organismo: vermelho para proteínas essenciais, cuja remoção levará à morte; laranja para proteínas de certa importância, cuja remoção irá retardar o crescimento; verde e amarelo para proteínas de menor importância ou de significado desconhecido.

Interações entre proteínas também têm contribuído para estudos de evolução das espécies. No mapeamento representado na figura 17, referente à rede proteica do verme *C. elegans*, Li et al. (2004) classificam as proteínas em três classes filogenéticas: (a) proteínas antigas, presentes em organismos unicelulares<sup>9</sup>, (b) proteínas mais novas, encontradas apenas em organismos multicelulares e (c) proteínas exclusivas do *C. elegans*. Longe de uma simples supremacia das proteínas novas em detrimento das mais antigas, as interações entre essas três classes filogenéticas ressoam um processo evolutivo complexo em que novas estruturas se constroem a partir das pré-existentes, formando com elas laços de interdependência:

---

<sup>9</sup> Mais especificamente trata-se de identificação de proteínas ortólogas, isto é proteínas presentes em diferentes espécies que embora não idênticas apresentam as mesmas funções.

Estes três grupos parecem ligar-se igualmente bem uns com os outros, sugerindo que as novas funções celulares dependem de uma combinação de elementos evolutivamente novos e antigos, em consonância com a clássica proposta da evolução atuando como um funileiro que para criar novas estruturas, modifica e faz acréscimos àquelas pré-existentes. (LI et al., 2004, tradução nossa).

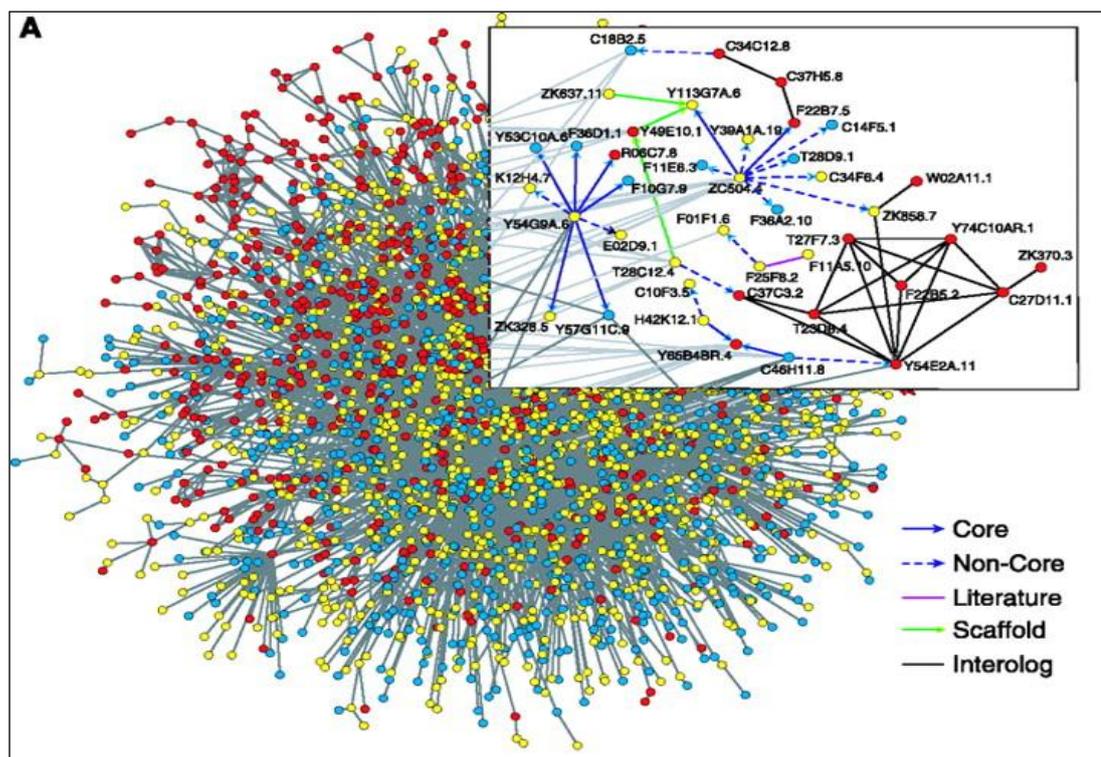


Figura 17 - Mapeamento das interações entre proteínas no *C. elegans* com as cores dos nós indicando sua classe filogenética (vermelho para as mais antigas, amarelo para as mais novas e azul para proteínas exclusivas do *C. elegans*) e cores de links indicando diferentes bases de dados.

As redes metabólicas constituem outro tipo de rede biológica em que a presença de *hubs* tem sido identificada. Neste caso, as reações químicas são tomadas como *links* entre as moléculas de reagentes e produtos que representam os nós da rede. De acordo com a investigação de Ma e Zeng (2003), o mapeamento de vias metabólicas como a da glicose (figura 18A) revela em diversos tipos de organismos uma distribuição de grau correspondente ao modelo da rede livre de escala (figura 18B). Além disso, verifica-se grande correspondência entre os principais *hubs* presentes em cada organismo, indicando que algumas poucas moléculas dominam a estrutura do conjunto das redes metabólicas (figura 18C).

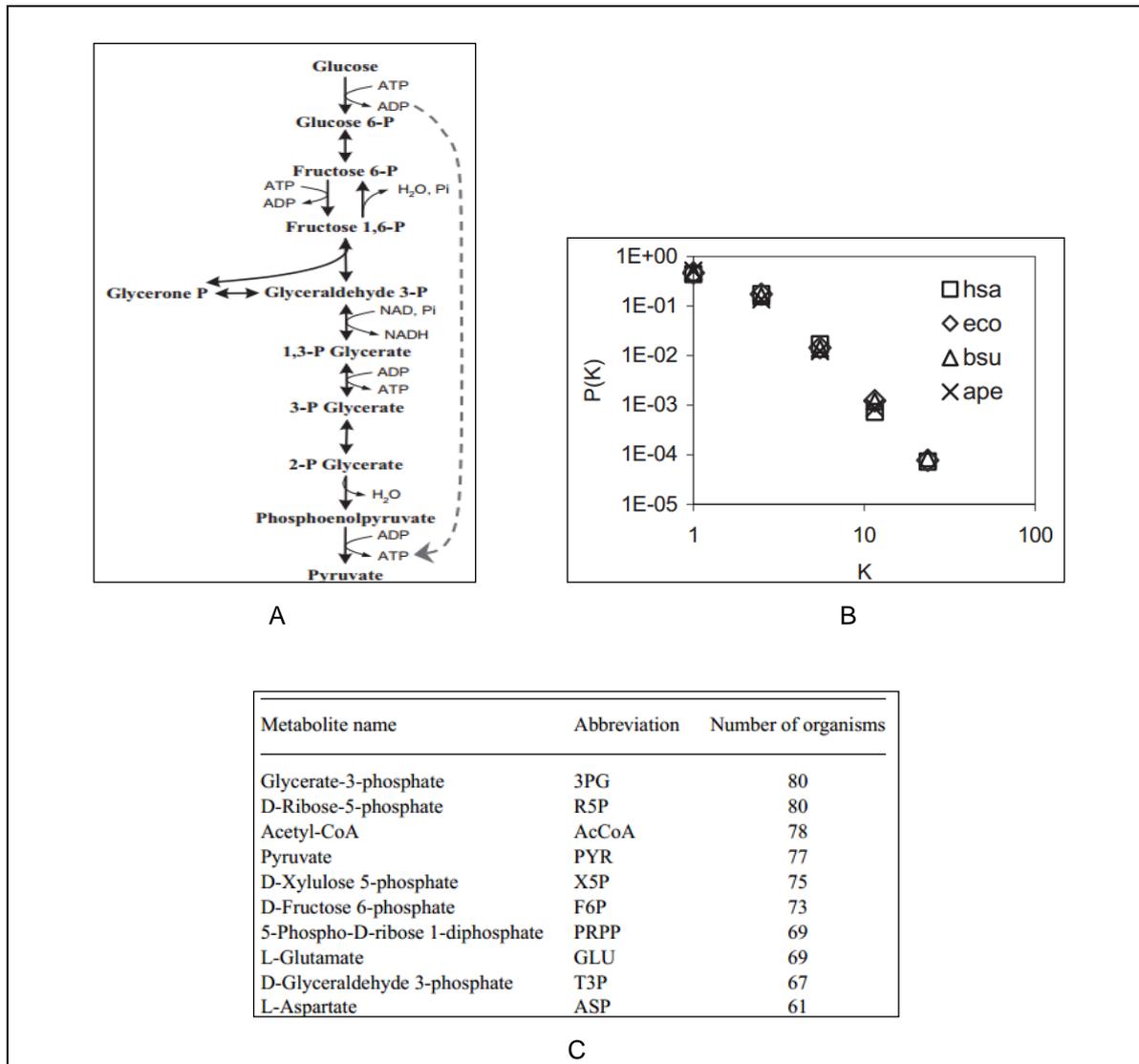


Figura 18 - (A) via metabólica da glicose; (B) distribuição de grau dos nós das redes metabólicas de vários organismos (hsa: *Homo sapiens* (eucarionte), eco: *Escherichia coli* (bactéria gram-negativa), bsu: *Bacillus subtilis* (bactéria gram-positiva), ape: *Aeropyrum pernix* (archaea)); (C) lista dos 10 metabolitos de maior grau.

Redes de transcrição genética com genes tomados como nós e suas ativações ou inibições por proteínas como *links*, redes de interação entre domínios proteicos, redes tróficas com nós constituídos por espécies e *links* pelas relações de predação e redes de relacionamentos entre animais de uma mesma espécie, como vespas e golfinhos, são outros exemplos de aplicação da teoria de redes na biologia.

### 3.3 REDES CEREBRAIS

O verme *C. elegans* tem cerca de 1 milímetro de comprimento e 70 micrômetros de espessura. Sem possuir visão ou audição, realiza o sensoriamento do ambiente através de receptores que respondem a estímulos táteis, químicos e térmicos. Possui apenas 302 neurônios, conectados e posicionados de maneira praticamente constante em todos os indivíduos da espécie. É em seu cérebro que se concentram a maioria destes neurônios, 279, responsáveis por 6393 *sinapses químicas*, 890 *sinapses elétricas* e 1410 *sinapses neuromusculares*<sup>10</sup>. A posição de cada neurônio e o comprimento de todas essas conexões de sinapses são bem conhecidas. Sabe-se, inclusive, que o conjunto das conexões entre tais neurônios apresenta características das chamadas redes de pequeno mundo e que alguns neurônios atuam como *hubs*.

Apesar de toda a precisão com que se conhece o mapa das células e conexões do sistema nervoso do *C. elegans*, ainda há um longo aprendizado a ser feito quanto às funções fisiológicas de cada neurônio. Modelagens computacionais como a desenvolvida pelo *Open Source Brain*<sup>11</sup> (figura 19) que procuram simular a transmissão de informações entre os neurônios desse animal ainda estão em seu estágio inicial de desenvolvimento.

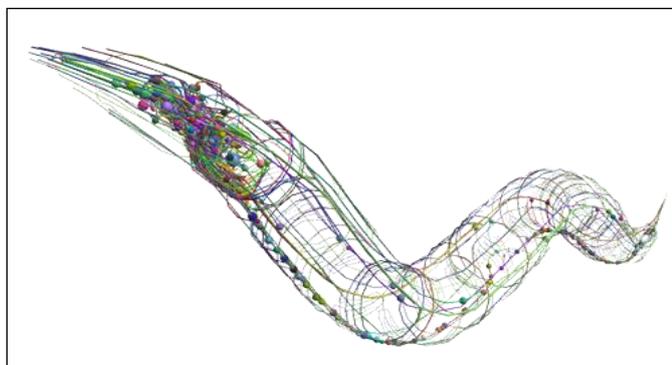


Figura 19 - Modelo elaborado pelo projeto *Open Source Brain* da anatomia celular tridimensional do *C. elegans* com destaque à sua rede de neurônios.

Se assim ocorre com este simples organismo, não é difícil imaginar as dificuldades presentes com relação ao estudo do cérebro humano que contém dez bilhões de células

<sup>10</sup> A sinapse, conexão de um neurônio (célula nervosa) com outra célula, pode ocorrer por intermédio da ação de substâncias químicas neurotransmissoras que funcionam como uma ponte por sobre a fenda existente entre as duas células (sinapse química) ou diretamente através de canais que unem as duas células que se apresentam quase justapostas (sinapse elétrica). Quando a conexão se dá entre um neurônio motor e uma célula muscular, a sinapse (química) é denominada de neuromuscular.

<sup>11</sup> Modelagem computacional acessível em <<http://www.opensourcebrain.org/projects/celegans#>> (Acesso em 31 Mar. 2014)

nervosas. A esta dificuldade estrutural soma-se a impossibilidade de realização de investigações anatômicas invasivas *in vivo* tal como acompanhar o traço deixado por substâncias injetadas no cérebro, técnica comumente realizada em ratos, gatos e macacos. Em vista da rápida deterioração do tecido nervoso após a morte, tal procedimento torna-se inútil para o estudo do cérebro humano.

De acordo com o texto de Sporns (2011), no qual em muito se baseia a exposição aqui feita sobre as redes cerebrais, há duas grandes vertentes técnicas para se mapear o cérebro humano: (a) investigação das conexões estruturais, anatômicas, entre neurônios fisicamente relacionados ou (b) investigação das conexões funcionais entre neurônios remotamente localizados. Estudos histológicos ou análises de imagens do tecido cerebral são técnicas comuns à identificação das conexões estruturais. Por sua vez, as conexões funcionais são inferidas por meio do acompanhamento da correlação entre séries temporais de dados, capturando padrões de desvios para com os dados estatísticos referentes ao funcionamento independente dos neurônios.

Seja qual for a vertente de investigação seguida, há ainda o problema prévio da definição dos nós e *links* a serem considerados. Além da construção da rede a partir das sinapses como *links* entre os neurônios, há várias outras possibilidades, tal como feito em experimentos eletrofisiológicos que consideram eletrodos e sensores como nós da rede. Como cada procedimento fornece pistas específicas e não excludentes, é comum o uso conjunto dos diversos tipos de mapeamento, tal qual indicado na figura 20.

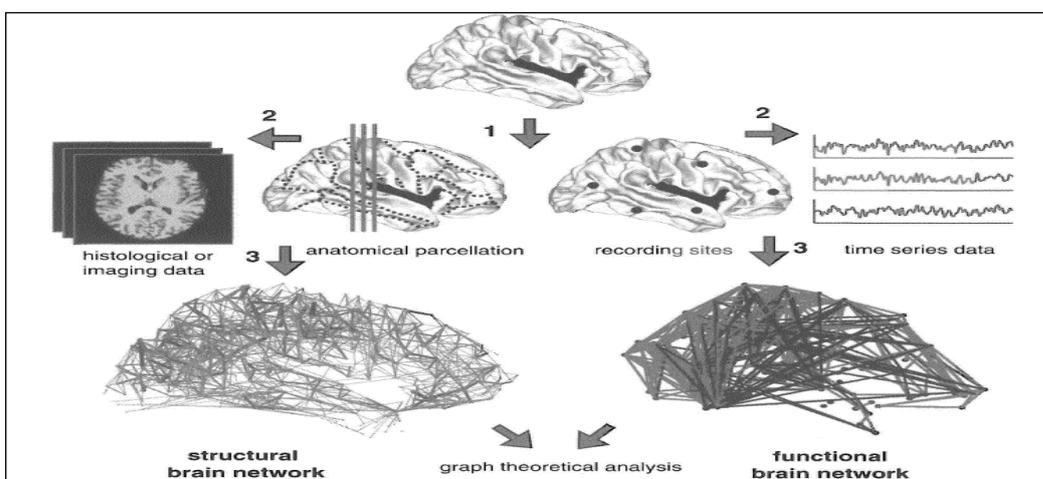


Figura 20 - Quatro principais passos no estudo das conexões estruturais e funcionais do cérebro humano: (1) definição dos nós por particionamento anatômico (à esquerda) ou por regiões funcionais (à direita); (2) identificação das conexões a partir da tomada de dados por imagens do tecido cerebral (à esquerda) ou por medidas da atividade cerebral (à direita); (3) geração da rede (estrutural à esquerda e funcional à direita) composta pelos nós e *links* identificados; (4) análise teórica e comparativa das duas redes obtidas. (SPORNS, 2011, p. 39, tradução nossa).

Fato é que apesar de hoje se conhecer muito mais do cérebro que há décadas atrás, sabe-se ainda muito pouco sobre tudo o que nele acontece. Segundo Sporns (2011), particular lacuna ocorre no terreno da relação entre a anatomia cerebral e a cognição:

Apesar de seu grande progresso a neurociência ainda não pôde responder às "grandes perguntas" sobre mente e inteligência. Consequentemente, a maioria dos cientistas da cognição continua a manter a posição de que a inteligência é fundamentalmente o trabalho de processamento simbólico realizado em arquiteturas baseadas em regras computacionais cuja função pode ser formalmente descrita de maneira totalmente independente de sua realização física. Se a cognição é majoritariamente simbólica na natureza, então seu substrato neural é pouco mais que um detalhe sem importância, nada revelando de essencial sobre a mente. Naturalmente, há muita controvérsia sobre o assunto. (SPORNS, 2011, p. 179, tradução nossa).

Entre o extremo do desprezo à rede neuronal ao reducionismo de que tudo a ela se resume, diversos experimentos têm utilizado técnicas computacionais integradas à teoria de redes para mapear alterações cerebrais em momentos de aprendizagem. Um exemplo é a pesquisa de Basset et al. (2011) acerca da reconfiguração dinâmica das redes cerebrais durante o aprendizado. Nessa investigação 18 pessoas, inseridas em uma máquina de ressonância nuclear magnética, responderam em 3 sessões diárias e ao longo de 5 dias a um teste referente a sequências de notas musicais.

As imagens relativas a cada indivíduo, escaneadas em regulares intervalos de tempo, foram particionadas nas 112 estruturas do córtex reconhecidas por atlas cerebral de referência. Tais estruturas foram, então, mapeadas como nós de uma rede conectados funcionalmente de acordo com o contraste entre a susceptibilidade magnética de hemoglobina oxigenada e desoxigenada dentro de cada elemento de volume das imagens de ressonância magnética. A figura 21 ilustra estas etapas do mapeamento de nós e *links* cujas propriedades permitiram aos pesquisadores constatar evidências de um aumento da organização funcional do cérebro durante o aprendizado.

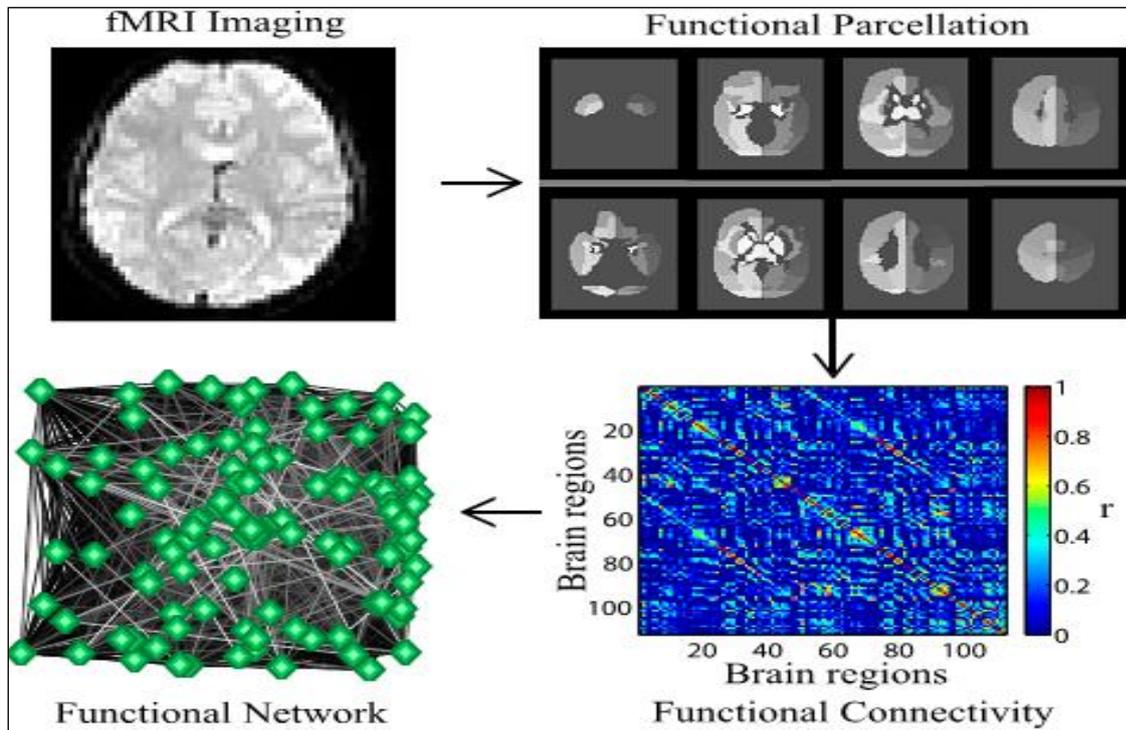


Figura 21 – Para gerar a rede funcional mapeada pelo experimento (*Funcional Network*), cada imagem de ressonância nuclear magnética (*fMRI Imaging*) foi particionada em conhecidas regiões do córtex (*Functional Parcellation*), consideradas, então como nós da rede correlacionados por uma matriz (*Functional Connectivity*) construída de acordo com a identificação dos sinais de contraste (BASSET et al, 2011).

### 3.4 INTERNET E WEB

Navegar pela *internet* não é sinônimo de navegar pela *world wide web* (*web*). A primeira é uma rede física de computadores e roteadores ligados por cabos elétricos, fibras ópticas e conexões sem fio. A segunda, constituída por documentos interconectados por *hiperlinks*, é apenas uma das diversas maneiras de trocar informações através da *internet*. Comunicação por *email*, *sms*, *tweets* e compartilhamento de arquivos são exemplos de outras redes de informações disponibilizadas pela *internet*, não necessariamente vinculadas à *web*.

A onipresença dos *microchips* tem inclusive possibilitado uma abrangência muito maior das conexões entre computadores, dando origem ao conceito de *internet de todas as coisas* (IoE, do inglês *internet of everything*). Segundo a CISCO (2013), hoje já seriam mais de 10 bilhões de pessoas, processos, dados e coisas (dispositivos móveis, parquímetros, termostatos, monitores cardíacos, pneus, estradas, carros, supermercados, etc.) conectadas via *internet*, com perspectiva de que esse número chegue a 50 bilhões em 2020.

Com tamanha heterogeneidade, abrangência e taxa de crescimento, já seria pertinente inferir a dificuldade inerente à tarefa de mapear a *internet*. Mas esse problema se amplia ainda mais quando se considera que a incorporação de novos nós e *links* (bem como a supressão de alguns mais antigos) é feita sem qualquer plano diretor geral e não raramente sujeita à emergência de tecnologias não previstas até mesmo por seus elaboradores. Nesse sentido, é ilustrativo o caso da invenção do *email* conforme relatada por Barabási (2002):

De fato, muitos usos atuais da internet seriam inimagináveis por aqueles que desenharam sua infraestrutura básica, a qual ainda está em vigor. Por exemplo, o email nasceu quando um aventureiro hacker, Rag Tomlinson, então trabalhando na BBN, uma pequena firma de consultoria em Cambridge, Massachusetts, descobriu como modificar o protocolo de transferência de arquivos para transportar mensagens de correio. Por um longo tempo, Tomlinson ficou quieto sobre seu avanço. Quando ele o expôs pela primeira vez a um de seus colegas, este o advertiu, “Não conte a ninguém! Não é sobre isto que nós deveríamos estar trabalhando”. O email vazou, no entanto, e tornou-se uma das aplicações dominantes do início da internet. (BARABÁSI, 2002, p. 149, tradução nossa).

Sendo da *internet* um recorte, se poderia supor maior facilidade nos estudos de mapeamento da *web*. A realidade, entretanto, não é essa, com cada um de seus usuários tendo acesso apenas a uma pequena parte de toda a rede. A razão, explica Barabási (2002), reside em que os *links* da *web* são direcionados:

Em outras palavras, ao longo de uma dada URL somente podemos viajar em uma direção. Se não houver link direto entre dois nós de uma rede direcionada, você pode conectá-los a partir de outros nós: por exemplo, se você quiser ir de A para D, você pode começar pelo nó A, então ir para o nó B, que tem um link para o nó C, que aponta para D. Mas você não pode fazer uma viagem de volta. Em uma rede não direcionada, em que você pode seguir um link em ambas as direções, um caminho  $A \rightarrow B \rightarrow C \rightarrow D$  implica que o caminho mais curto de D para A, é o caminho reverso  $D \rightarrow C \rightarrow B \rightarrow A$ . Em uma rede direcionada, no entanto, não há garantia de que exista o caminho inverso. O mais provável é que você terá que seguir uma rota diferente em seu caminho de volta: partindo de D, talvez você precise visitar dúzias de nós intermediários antes de voltar para A. A web é repleta de tais caminhos desarticulados. Eles fundamentalmente determinam a navegabilidade da web. (BARABÁSI, 2002, p. 167, tradução nossa).

De acordo com o estudo de Andrei Broder et al. (2000) este direcionamento dos *links* da *web* a tornam fragmentada em quatro principais continentes: (a) núcleo central, em que todas as páginas podem alcançar umas às outras por um *link* direto entre elas ou por uma sucessão de *links* direcionados; (b) páginas de acesso, contendo páginas que podem chegar ao núcleo central, mas não podem ser alcançadas a partir dele, como sites recentemente inaugurados e ainda não descobertos pelos usuários; (c) páginas de saída, com páginas acessíveis a partir do núcleo central, mas sem ligação de volta a ele, como sites corporativos que contêm apenas ligações internas; (d) tentáculos, que contêm páginas que não podem

chegar ao núcleo central e nem serem alcançadas a partir dele. Há, ainda, páginas de acesso que podem se conectar diretamente a páginas de saída (tubos) e conjuntos de páginas completamente desconectadas de todo o resto (ilhas). A figura 22, apresentada por Barabási (2002), ilustra bem essa visão da *web*.

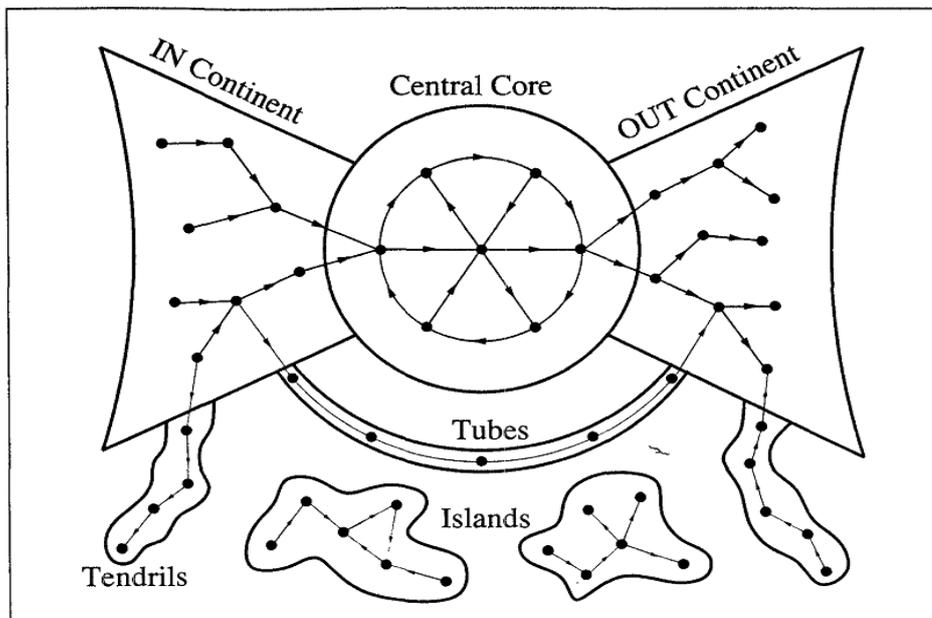


Figura 22 – A fragmentação da *web* contendo núcleo central (*Central Core*), páginas de acesso (*IN Continent*), páginas de saída (*OUT Continent*), tentáculos (*Tendrils*), tubos (*Tubes*) e ilhas (*Islands*). (BARABÁSI, 2002, p. 166).

Tão relevante quanto a estrutura fragmentada, é o equilíbrio existente na distribuição dos nós entre estas regiões. Dos cerca de 203 milhões de nós mapeados pelo *software* de rastreamento utilizado por Broder, apenas 56 milhões pertencem ao núcleo central, com as páginas de acesso e de saída contendo cerca de 43 milhões cada e os 61 milhões restantes sendo internos às ilhas e tentáculos. Isso significa que mesmo que determinado *website* seja um robusto *hub* do núcleo central, metade da *web* é para ele invisível.

Importante notar que tal fracionamento é independente de regras regulatórias, poder econômico, *marketing* ou técnicas computacionais mais eficientes. Trata-se de uma propriedade topológica da *web* e de todas as redes formadas por *links* direcionados, haja vista as redes tróficas em que predadores se alimentam de suas presas, sem validade para o caminho inverso.

### 3.5 EMPRESAS EM REDE

Um dos aspectos marcantes da sociedade contemporânea, conforme caracterizada por Castells (2009), é a emergência das empresas em rede, nova forma de organização dos negócios em que diferentes firmas, segmentos de firmas ou as próprias firmas em sua constituição interna se estruturam em redes ao mesmo tempo competitivas e colaborativas:

Assim, grandes empresas são internamente descentralizadas como redes. Pequenas e médias empresas são conectadas em redes, garantindo a massa crítica de suas contribuições como subempreiteiras, mantendo seu principal ativo: a flexibilidade. Redes de pequenas e médias empresas são muitas vezes complementares para grandes empresas; na maioria das vezes para várias delas. Grandes corporações, e suas redes subsidiárias, costumam formar redes de cooperação, denominada, na prática de negócios, alianças estratégicas ou parcerias.

Mas essas alianças são estruturas cooperativas raramente permanentes. Este não é um processo de cartelização oligopolista. Essas redes complexas se articulam em projetos de negócios específicos e reconfiguram sua cooperação em diferentes redes a cada novo projeto. A prática comercial habitual nesta economia em rede é a das alianças, parcerias e colaborações específicas para determinado produto, processo, tempo e espaço. A fim de ganhar maior participação no mercado, estas colaborações são baseadas em partilha de capital, trabalho e fundamentalmente, de informação e conhecimento. Então, elas são principalmente redes de informação, que ligam os fornecedores e clientes através da empresa em rede. A unidade do processo de produção não é a empresa, mas o projeto do negócio, promulgado por uma rede, a empresa em rede. A empresa continua a ser a unidade legal de acumulação de capital. Mas, desde que o valor da empresa em última análise, depende da avaliação financeira no mercado de ações, a unidade de acumulação de capital, a empresa, torna-se um nó da rede global de fluxos financeiros. (CASTELLS, 2009, p.32, tradução nossa).

Boa percepção desse enredamento empresarial pode ser obtida pela análise da rede das corporações em multimídia, composta por um núcleo central de empresas gigantes conectadas tanto entre si (*Time Warner, Disney, News Corporation, Bertelsmann, NBC Universal, Viacom e CBS*), como com empresas gigantes da *internet (Google, Yahoo, Apple, Microsoft)* e com diversas firmas de menor porte.

A figura 23 ilustra bem a rede dos negócios de multimídia. Enquanto a francesa *Vivendi* tem uma participação de 20% da *NBC* norte-americana, esta detém 25% da também norte-americana *A&E Networks*, cujos demais 75% apresentam-se distribuídos entre as igualmente norte-americanas *Hearst* e *Walt Disney*, com esta última também detendo 20% da *TVA* brasileira e tendo 1% de seu capital sob domínio da empresa árabe *Kingdom Holdings* que detém também 1% da norte-americana *Time Warner*, proprietária de 24% da indiana *Zee Turner Limited*, cujos restantes 74% estão em mãos da indiana *Zee Telefilms*.

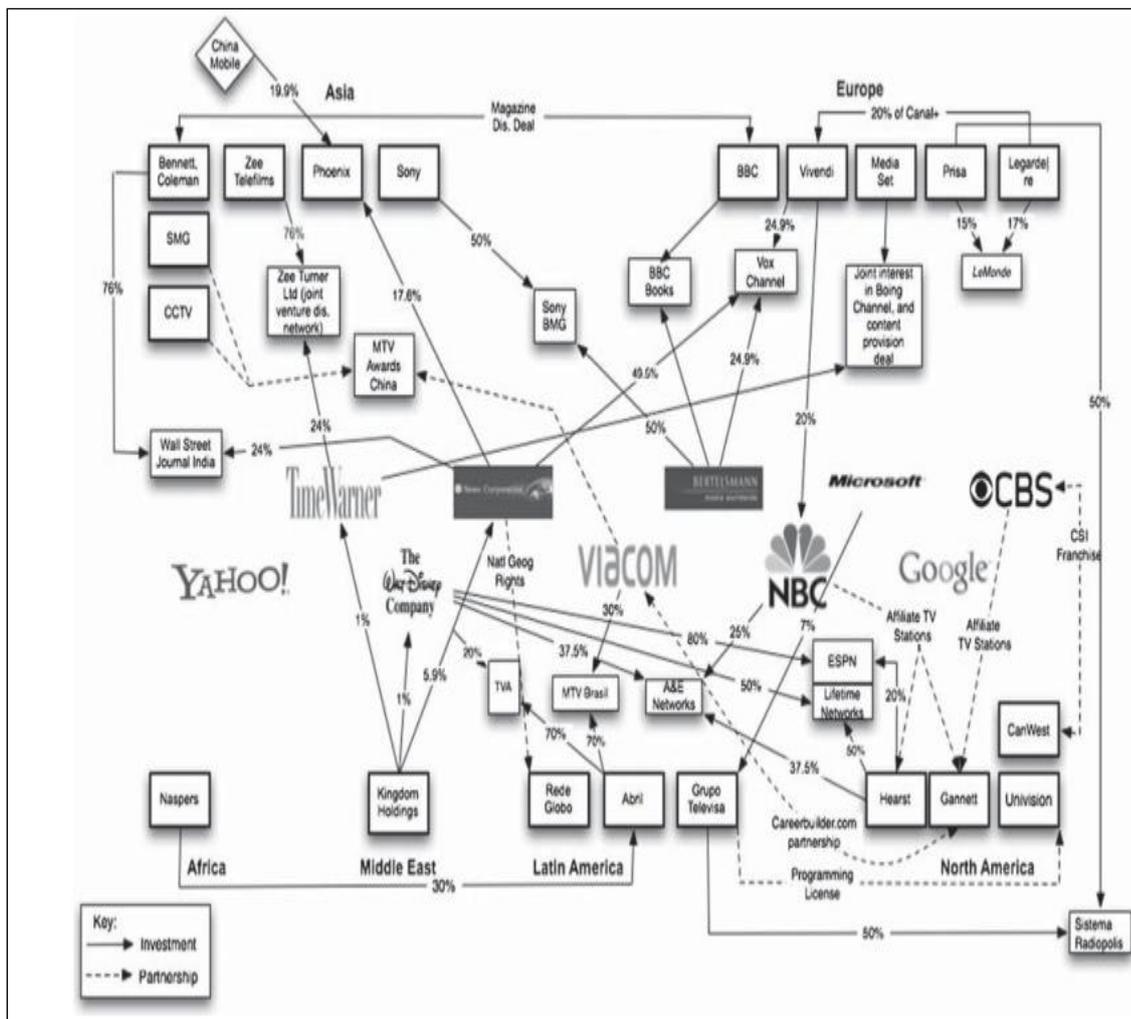


Figura 23 – Mapeamento de rede entre empresas de multimídia (CASTELLS, 2009, p. 85).

Segundo Castells (2009) nesse emaranhado intercontinental, tanto o global influencia o local como o local influencia o global. Exemplo do primeiro fenômeno é a exibição em todo o mundo de canais de TV como HBO, ESPN, FOX, e CNN, bem como de programas como *The Voice*, *Pop Idol*, *Survivor* e *Who Wants to be a Millionaire*. No entanto, ao mesmo tempo em que corporações globais introduzem seus produtos em rede ao redor do mundo, empresas locais utilizam essa mesma rede para inserir seus produtos em outras regiões, produtos esses que por vezes tornam-se globais: a história e personagens de *Rei Leão* baseiam-se no mangá japonês, ao passo que o *Big Brother* tem origem alemã. Conforme analisado por Castells (2009), a globalização da produção e do capital e as identidades locais ou nacionais são, enfim, inter-relacionadas:

Um pequeno número de megacorporações constituem a espinha dorsal da rede global das redes de mídia. O seu domínio se baseia em sua capacidade de influenciar

e se conectar, em todos os lugares, às organizações de mídia centradas local e nacionalmente.

Por outro lado, a fim de facilitar sua própria expansão, organizações de mídia com centro nacional e regional dependem cada vez mais das parcerias com essas megacorporações. Embora o capital e a produção sejam globalizados, o conteúdo da mídia é personalizado para as culturas locais e para a diversidade de públicos segmentados. Assim, tal como é típico de outras indústrias, a globalização e a diversificação trabalham lado a lado. Na verdade, os dois processos estão interligados: apenas redes globais podem dominar os recursos de produção da mídia global, mas sua capacidade para conquistar participação no mercado depende da adaptação do seu conteúdo ao gosto do público local. O capital é global; identidades são locais ou nacionais. (Castells, 2009, p.72, tradução nossa).

### 3.6 COAUTORIAS

De que maneira pesquisas realizadas ao redor do mundo acerca de determinado ramo científico, como a nanotecnologia, se entrelaçam? Quais países apresentam maior número de trabalhos publicados sobre esse assunto e como colaboram entre si os pesquisadores de diferentes países? Se tais pesquisadores forem agrupados pelas instituições a que pertencem, como se apresenta o mapa da rede de colaborações institucionais sobre nanotecnologia?

Perguntas como essa podem agora ser respondidas com relativa facilidade através de pesquisas em extensas bases de dados de índices de artigos científicos como as disponibilizadas pelo projeto *Web of Science* (WoS)<sup>12</sup>. Assim procedeu o Observatorio Iberoamericano de la Ciencia, la Tecnología y la Sociedad (OBSERVATORIO CTS, 2011) que a partir da identificação de cerca de 300 mil publicações entre 2000 e 2007 relativas à nanotecnologia, realizou diversas análises sobre a produção científica neste campo.

A figura 24 representa um destes estudos, em que os coautores dos artigos científicos publicados em 2007 foram agrupados pelos países de suas instituições. Trabalhos publicados em conjunto por pesquisadores de uma instituição brasileira e outra norte-americana, por exemplo, configuram nesta rede um *link* entre Brasil e Estados Unidos. Dada a grande quantidade de nós e *links*, na elaboração desta representação visual em rede, o Observatorio CTS (2011) promoveu alguns cortes, de forma a eliminar os *links* menos importantes, deixando tão somente a quantidade mínima necessária para que nenhum nó aparecesse desconectado. Além disso, foram considerados apenas os países com pelo menos 10 publicações naquele ano.

---

<sup>12</sup> Base de dados acessível mediante acesso institucional (USP) em <<http://wokinfo.com/>> (Acesso em 31 Mar. 2014).

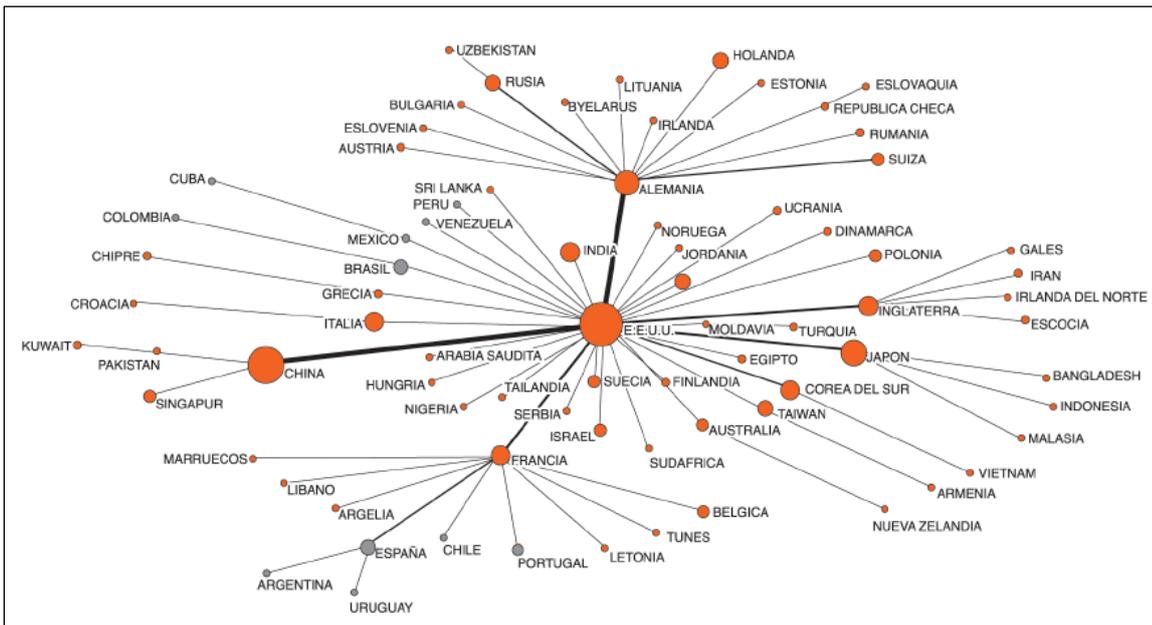


Figura 24 - Rede de países com produção científica em nanotecnologia. Mapeamento elaborado pelo Observatório CTS (2011) a partir de dados da *WoS*, incluindo-se somente publicações de 2007 de países com mais de 10 artigos, sendo os *links* mais espessos indicativos de maior quantidade de trabalhos elaborados em conjunto, o maior diâmetro do nó correspondendo à maior quantidade de artigos publicados pelo conjunto dos pesquisadores do país e a cor cinza representando os países ibero-americanos.

Comparando com as redes relativas a cada ano no período de 2000 a 2007, o Observatório CTS (2011) conclui que a trama de relações entre os países cresceu em complexidade, sempre mantendo os Estados Unidos como nó principal, tanto pela quantidade de publicações, como por ser o nó ao redor do qual se articulam os demais países, mas com notável fortalecimento do poder aglutinador de outros países, como Alemanha e França de onde se expandem ramos secundários da rede.

Reagrupando os artigos a partir das instituições de origem dos coautores e focando apenas nas relações entre as vinte e cinco instituições ibero-americanas de maior produção científica em nanotecnologia, nova representação foi elaborada (figura 25), na qual se percebe a preponderância do Consejo Superior de Investigaciones Científicas (CSIC), espanhol, não apenas para a própria Espanha, como também para a integridade de toda a rede regional. Quanto às instituições brasileiras se percebe que estão fortemente conectadas tanto entre si como internacionalmente, destacando-se a Universidade de São Paulo (USP) pela quantidade de publicações e a Universidade de Campinas (UNICAMP) com sua relação direta com mais da metade das entidades representadas na rede. São, aliás, estes vínculos internacionais que conferem à UNICAMP posicionamento mais ao centro da figura do que a própria USP ou do que a Universidad Autónoma de México (UNAM) e a Universidad de Barcelona, todas com

maior volume de publicação do que ela. Tal efeito se deve ao uso do *Kamada-Kawai*, um dos algoritmos que utiliza o já comentado modelo de massas (nós) e molas (*links*) configurado no estado de menor energia possível (ver página 57), promovendo que o centro da figura tenda a ser ocupado por nós que além de apresentarem muitas conexões, sejam fortemente conectados entre si: a USP ocupa região central na rede estabelecida entre as universidades brasileiras, mas ao ter menos trabalhos sobre nanotecnologia em coautoria com instituições de outros países ibero-americanos, tem posição mais periférica do que a UNICAMP nessa rede internacional.

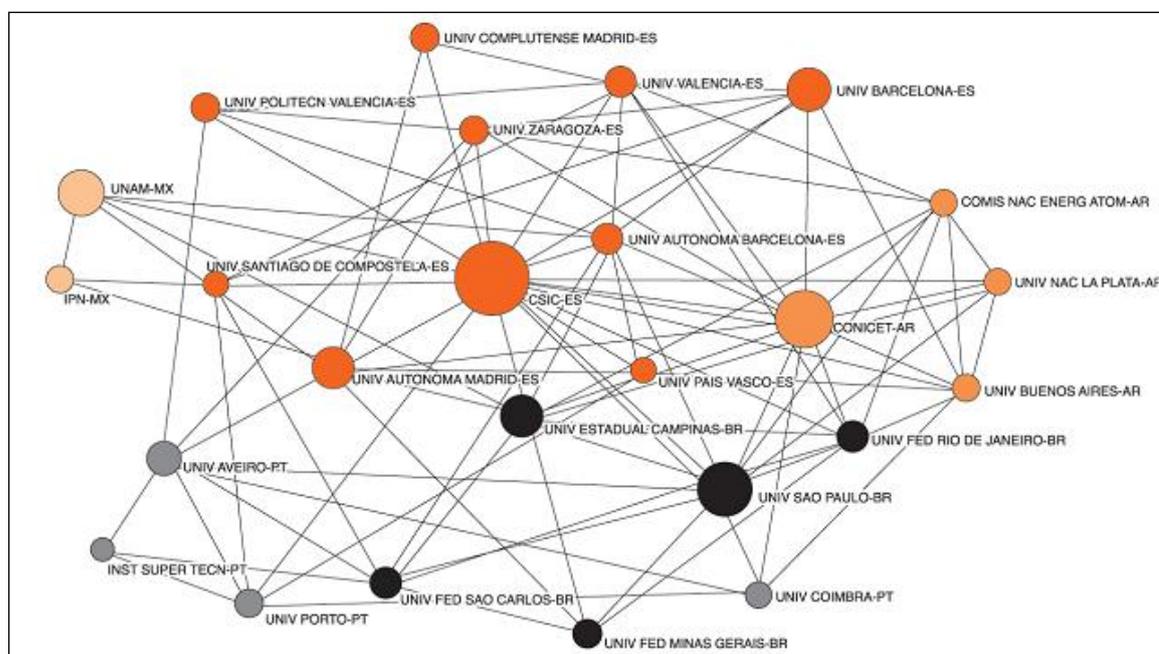


Figura 25 - Rede de instituições ibero-americanas. Mapeamento elaborado pelo Observatório CTS (2011) a partir de dados da *WoS*, incluindo-se somente as 25 instituições ibero-americanas de maior produção científica em nanotecnologia em 2007, sendo os *links* mais espessos indicativos de maior quantidade de trabalhos elaborados em conjunto, o maior diâmetro do nó correspondendo à maior quantidade de artigos publicados pelo conjunto dos pesquisadores da instituição e com as cores dos nós representando os diferentes países (laranja mais forte para Espanha, preto para Brasil, cinza para Portugal, laranja muito claro para México e laranja claro para Argentina).

### 3.7 COOCORRÊNCIA DE CITAÇÕES

Além da análise de coautorias, os dados disponibilizados pela *WoS* também permitem mapear redes de citações como a representada na figura 26 em que se identificam os dezoito trabalhos indicados como referências bibliográficas no famoso artigo em que César Lattes e demais autores descrevem o *méson pi* e os cento e vinte e quatro artigos que o utilizam como referência.



Contando com a mesma técnica de simplificação (eliminação dos *links* menos importantes sem deixar nenhum nó desconectado), esta rede conta com cento e seis disciplinas com destaque à química, ciência dos materiais, física, ciência dos polímeros, engenharia, bioquímica e biologia molecular, com o forte vínculo entre as duas últimas resultando em acoplamento de nó único. Três blocos temáticos são evidenciados na análise Observatorio CTS (2011): (a) a química e suas densas conexões com a ciência dos materiais, com a ciência dos polímeros e com a bioquímica e biologia celular; (b) a ciência dos materiais e suas ligações com a física e a engenharia; (c) a bioquímica e a biologia molecular contendo ramificações que abrangem grande número de disciplinas.

Recortes por nacionalidades permitiram ao Observatorio CTS (2011) tecer comparações entre a rede geral internacional com as redes de disciplinas relativas a cada país ibero-americano, bem como ao conjunto dos países da Ibero-América. O que se percebe é que embora o quadro ibero-americano não apresente significativas diferenças para com o conjunto internacional, a análise em detalhe, país a país, permite identificar algumas importantes particularidades. Um exemplo é o surgimento na rede de nanotecnologia brasileira, composta por 71 disciplinas, de uma ramificação a partir da oncologia para a medicina, radiologia e otorrinolaringologia (figura 28). O Observatorio CTS (2011) infere que tal especificidade deva-se a uma maior preocupação dos pesquisadores brasileiros para com investigações em nanotecnologia vinculadas às ciências da saúde.

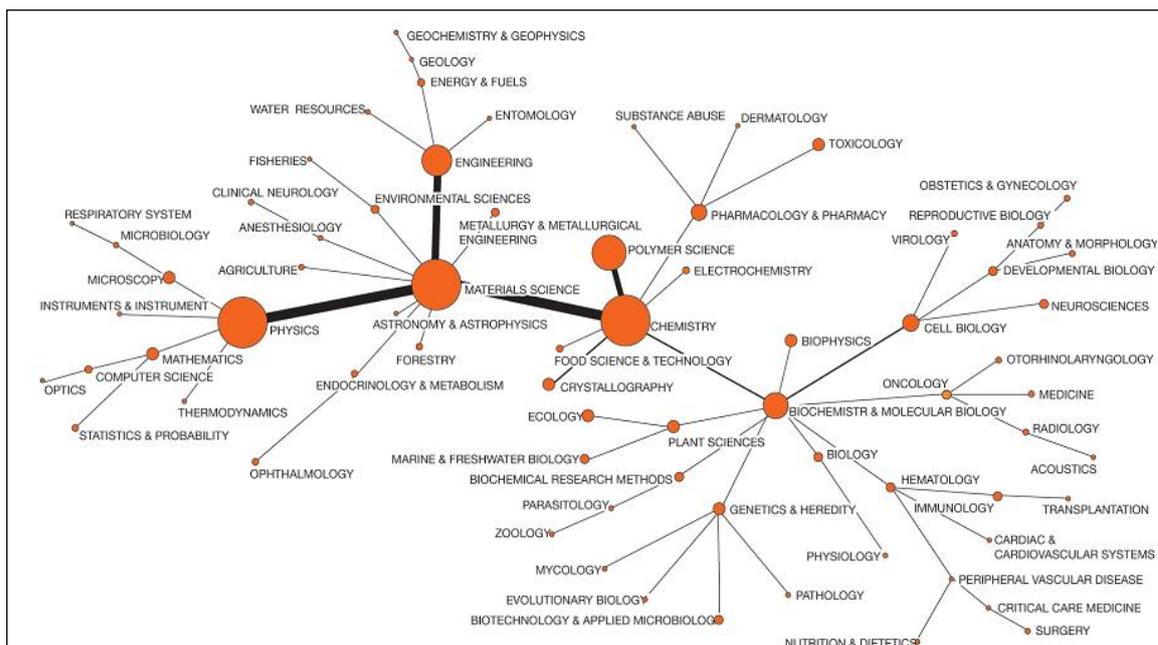


Figura 28 - Rede de áreas científicas citadas em trabalhos brasileiros de nanotecnologia.

Procurando contribuir com o mapeamento das pesquisas e publicações realizadas no Brasil em *Ciência, Tecnologia e Sociedade* (CTS), Chrispino et al. (2013) procuraram em 22 periódicos científicos brasileiros da área de ensino, no período entre 1996 e 2010, por artigos que continham as palavras-chave ciência, tecnologia e sociedade, separadas ou juntas, tendo encontrado após filtragem de atenta leitura 88 artigos efetivamente relativos à área e que lhes forneceram 394 fontes bibliográficas. Com esses dados em mãos, construiu-se uma rede direcionada composta pelas fontes bibliográficas como nós que se interligavam pelas suas correspondências: se uma fonte bibliográfica A citava outra B, ficava estabelecido um *link* de A para B. Após refinamento dos dados obtidos, eliminando-se os nós não conectados ao principal componente, chegou-se à rede apresentada pela figura 29, organizada através das partições de nós de mesmo grau e com a identificação dos 13 trabalhos mais citados.

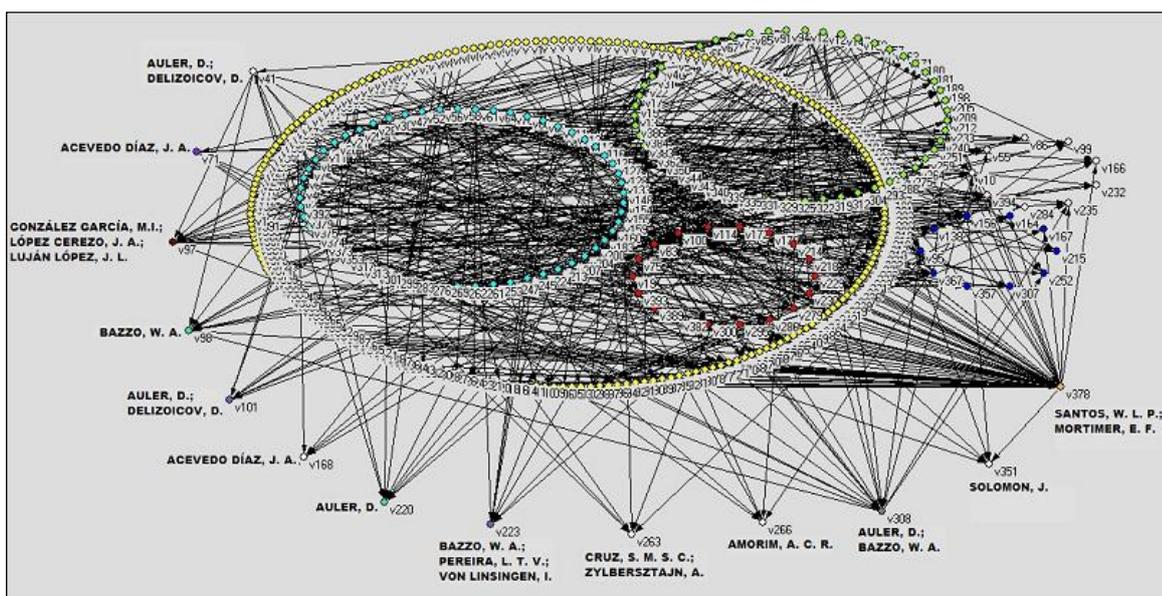


Figura 29 - Mapeamento das pesquisas e publicações realizadas no Brasil em Ciência, Tecnologia e Sociedade (CTS), conforme conexões identificadas a partir de artigos publicados em 22 periódicos, com destaque aos 13 trabalhos mais citados e com vértices de cores iguais possuindo o mesmo grau de entrada.

Em vista dos resultados, os autores sugerem que estes 13 trabalhos “*podem ser encarados como leitura obrigatória para entendimento do que seja e o que produz a área. É o ponto comum, é o consenso, se não construído intencionalmente, certamente estabelecido pela tradição*” (CHRISPINO et al., 2013). Além dessa afirmação, são tecidas diversas inferências possibilitadas por essa visão em rede dos trabalhos em educação CTS no Brasil, tais como a predominância de autores nacionais como fonte de pesquisa, a ausência de citações a autores tidos por historiadores como fundadores da área e a constatação de

tendências endógenas que contrastam com a interdisciplinaridade e a contextualização características da área CTS.

### 3.8 MINERAÇÃO DE TEXTOS E REDES CONCEITUAIS

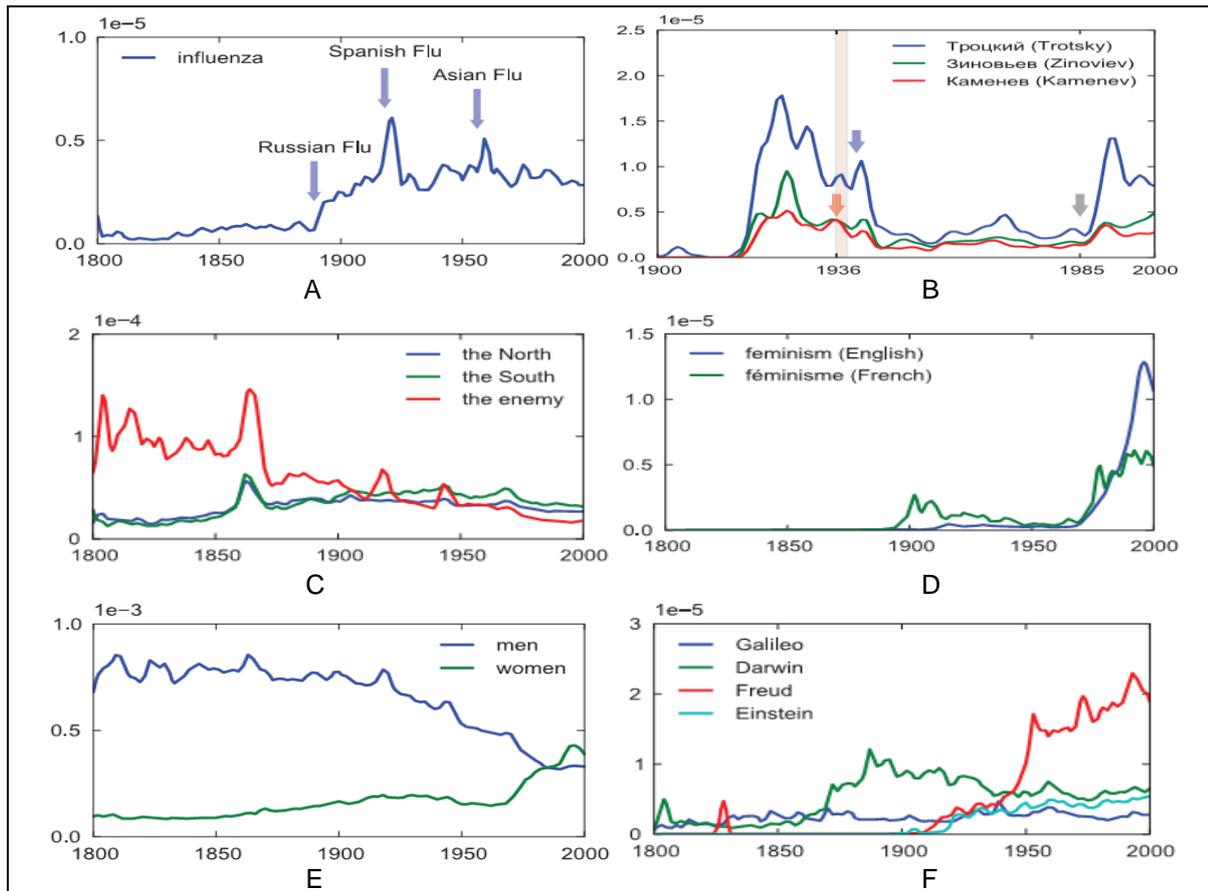
Nas elaborações das redes de citações apresentadas no item anterior, o primeiro passo foi a coleta dos documentos referentes à temática em foco, esforço realizado através do “garimpo” de termos relevantes em textos contidos em uma determinada base de dados. De maneira semelhante, tornou-se comum a realização de pesquisas na *web*, uma imensa base de dados que junto com suas facilidades traz crescentes preocupações quanto à originalidade de textos apresentados para efeito de trabalhos escolares ou mesmo científicos. A mesma *web* apresenta, no entanto, antídotos contra tais fraudes, com alguns professores e orientadores adotando a prática de “perguntarem ao Google” se essa ou aquela frase já não se encontra presente em algum hipertexto.

Google, aliás, que disponibiliza o *Google Books Ngram Viewer*<sup>13</sup>, uma base de dados construída com a digitalização de milhões de livros, em sua maioria na língua inglesa, que possibilita pesquisas sobre a evolução de palavras e expressões ao longo do tempo. Mais que isso, de acordo com Michel et al. (2010), os dados fornecidos pelo projeto permitem o estudo de diversos fenômenos culturais refletidos no uso da linguagem, motivando o reconhecimento de um novo campo de pesquisa denominado pelos autores de “*culturomics*” cujos resultados constituem “*um novo tipo de evidências em ciências humanas*” e, “*tal como acontece com fósseis de criaturas antigas, o desafio de culturomics reside na interpretação dessa evidência*”(tradução nossa).

Como exemplo de tais evidências, a figura 30 apresenta alguns dos vários gráficos elaborados pelos autores referentes a evoluções históricas e culturais: (a) oscilação do uso da palavra “*influenza*”, com picos correspondentes às conhecidas pandemias na Rússia, Espanha e Ásia; (b) supressão de Trostky, Zinoviev e Kamenev nos textos russos durante a ascensão e desenvolvimento do *stalinismo*; (c) trajetória histórica das expressões “*the North*”, “*the South*” e “*the enemy*”, indicativa da polarização entre estados durante a guerra civil norte-americana; (d) a ascensão do feminismo primeiramente na França e depois, com mais vigor nos Estados Unidos; (e) crescente ganho de espaço da mulher frente ao homem na “*batalha dos sexos*” (f) a maior popularidade de Freud frente a Galileu, Darwin e Einstein.

---

<sup>13</sup> Base de dados disponível em <<https://books.google.com/ngrams/>> (Acesso em 31 Mar. 2014).



A investigação das frequências com que as palavras são utilizadas em textos podem também revelar características de seus autores. Ao expor um método de extração automática de termos e conceitos de textos que caracterizem a ontologia de campos de conhecimento específicos, Ahmad e Gillam (2005) argumentam que os especialistas de cada área tendem a repetir selecionadas palavras-chave a fim de consolidar ou rejeitar determinados conceitos, desenvolvendo, assim, linguagem própria da área. No caso da física nuclear, os autores relatam que a palavra partícula é, por exemplo, cerca de seiscentas vezes mais frequente nos textos de Rutherford do que na coletânea disponibilizada pelo *British Nacional Corpus* (BNC)<sup>14</sup>, banco de dados contendo mais de 100 milhões de palavras extraídas de diversas fontes e que busca representar mostra significativa do inglês britânico atualmente falado e escrito. Estendendo a análise para outras palavras frequentemente utilizadas pelo próprio

<sup>14</sup> Base de dados disponível em <<http://www.natcorp.ox.ac.uk/>> (Acesso em 31 Mar. 2014).

Rutherford, Bohr ou modernos físicos nucleares, Ahmad e Gilliam (2005) analisam que a desproporcionalidade no uso de determinadas palavras fornece uma pista do compromisso do especialista para com uma determinada concepção ou modelo teórico:

As mais frequentes palavras em Rutherford incluem *partícula(s)*, *átomos* e *núcleo*. Estas palavras são “desproporcionalmente” usadas por Rutherford quando comparadas com um texto típico em inglês – ele usa *partícula* 629 vezes mais frequentemente do que usado no British National Corpus, *átomo* 896 vezes mais frequentemente e *núcleo* 841 vezes mais frequentemente. Aqui há indícios dos famosos experimentos de espalhamento – onde Rutherford mediu o *alcance* (22 vezes mais frequente) das partículas *alfa* (485 vezes mais frequente) emitidas por uma fonte radioativa (em *centímetros* ou *cm*). A ênfase em Bohr é sobre os *elétrons* (1652 vezes mais frequente) e suas *órbitas* (1204 vezes mais frequente); *núcleo* é usado com menor desproporção em Bohr (652 vezes) do que em Rutherford (841 vezes). O uso mais frequente dos *elétrons* em uma *órbita* não minimiza o fato de que a órbita era ao redor do *núcleo*. A palavra *núcleon* (um hipônimo para *próton* e *nêutron*) está entre as mais desproporcionalmente utilizadas [na seleção de textos de modernos físicos nucleares] – 36410 vezes mais frequente em nossos textos do que no BNC; *energia* (e sua unidade *mev* – milhões de eletron-volts) está entre as mais frequentemente utilizadas. A frequência de [uso das palavras] *transversal*, *seção*, e *espalhamento*, reflete o uso do termo *seção-transversal* em física nuclear onde é usado para se referir à medida da probabilidade de uma reação *nuclear*; *seção transversal de espalhamento* é [um termo] usado na determinação da estrutura do núcleo. (Ahmad e Gillam, 2005, tradução nossa preservando o uso original do estilo itálico como indicação de palavra frequentemente usada em cada um dos três conjuntos analisados e acrescentando o estilo sublinhado como indicativo do início da análise de cada um desses conjuntos).

Se a intenção não for a investigação da frequência com que uma palavra apareça em diversos textos, mas como se distribuem as palavras em determinado texto, há diversos softwares que automatizam a tarefa. Um deles é o *Wordle*<sup>15</sup> que após realizar a contagem das diferentes palavras dispostas no texto, valorar cada uma delas proporcionalmente à sua frequência e eliminar palavras de ligação como artigos e preposições, desenha uma nuvem de palavras disposta horizontal e verticalmente. A figura 31 ilustra a nuvem de palavras construída pelo *Wordle* para os dois primeiros parágrafos daquele texto fictício do início do dia de uma secretária apresentado na página 38 deste trabalho. Ótima para permitir rápida percepção do assunto tratado no texto, a automatização do processo não permite a interferência na lista das palavras a serem eliminadas, implicando em equivalente destaque para conceitos, como “cristal”, “tempo”, “sinal” e “TV”, e palavras comuns, tal qual a preposição “sobre”. Se, por um lado, a fim de melhor identificar o tema tratado no texto, essa equivalência pode ser indesejada, por outro lado, ela pode denunciar a presença de vícios de linguagem e inferir possíveis usos didático-pedagógicos da ferramenta no aprendizado da redação de textos.

---

<sup>15</sup> Ferramenta disponível em <<http://www.wordle.net/>> (Acesso em 31 Mar. 2014).



Entre a disposição do texto como nuvem de palavras e como rede que enlaça as palavras adjacentes, têm-se desenvolvido procedimentos intermediários que apontam para a construção de redes entre conceitos, termos ou palavras-chave. A ideia básica é proceder à representação em rede somente após uma filtragem no texto, seja com a eliminação das palavras de ligação e de outros termos comumente utilizados em textos não específicos, seja com a “*lematização*” das palavras, isto é, agrupamento das palavras apresentadas em suas diversas derivações, tais como tempo verbal, gênero e número (“*explicação*”, “*explicado*”, “*explicadas*”, por exemplo, podem ser agrupadas como derivações de “*explicar*”).

Abordagem nessa direção foi apresentada por Brunn (2012) em sua análise da compreensão de conceitos e fenômenos científicos por parte de estudantes ao responderem a problemas elaborados por seus professores. A hipótese é que a conversão do texto do aprendiz em uma rede de associação linguística “reduzida”, condicionada pela transformação das palavras comuns à própria língua escrita em *links* entre as palavras adjacentes restantes (figura 33), permita que estruturas mentais presentes em segundo plano no texto ganhem maior destaque, auxiliando a avaliação das competências do estudante. Segundo Brunn (2012), redes reduzidas que tenham maior número de nós e maior fluxo de informações<sup>16</sup> são provavelmente indicativas da maior habilidade do estudante em utilizar a física em diferentes situações.

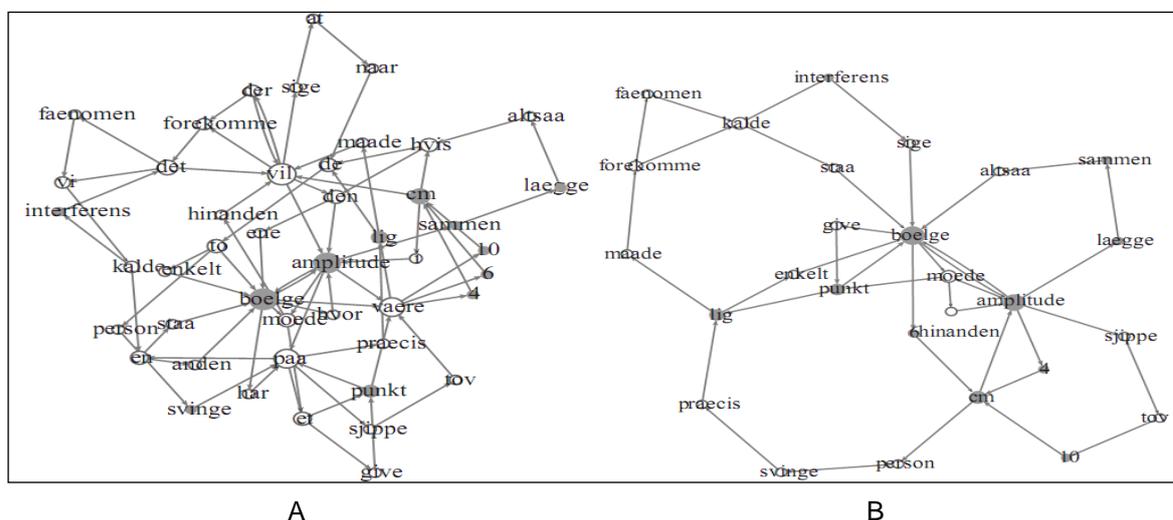


Figura 33 - Duas redes linguísticas construídas por Brunn (2012) com base à “*lematização*” de palavras: (A) sem eliminação das palavras de ligação; (B) com eliminação das palavras comuns do idioma dinamarquês. Em cinza, nós associados pelo autor ao fenômeno ondulatório da interferência.

<sup>16</sup> Por fluxo de informações entenda-se as medidas de duas grandezas utilizadas por Brunn: a “*informação de busca*” (*search information*), associada ao número de *caminhos geodésicos* entre dois nós, e a entropia, associada ao número de *caminhos geodésicos* para se alcançar determinado nó. Quanto maior a informação de busca de uma rede, maior sua habilidade em localizar um nó B a partir de um nó A. Quanto maior a entropia de uma rede, menor a previsibilidade de como determinado nó pode ser alcançado. Assim, baixos valores para a informação de busca associados a altos valores de entropia são indicativos de facilitação no fluxo de informações dentro da rede.

Também utilizando a técnica de *lematização* e eliminação das palavras de ligação, Antiqueira et al. (2005) representaram textos como redes de ligação entre palavras adjacentes, com o diferencial de associarem pesos a cada *link*, de maneira que quanto maior o número de vezes que as mesmas duas palavras se apresentassem adjacentes, maior o valor atribuído ao respectivo *link*. A figura 34 ilustra um exemplo do processamento realizado pelos autores referente a um poema de Carlos Drummond de Andrade.

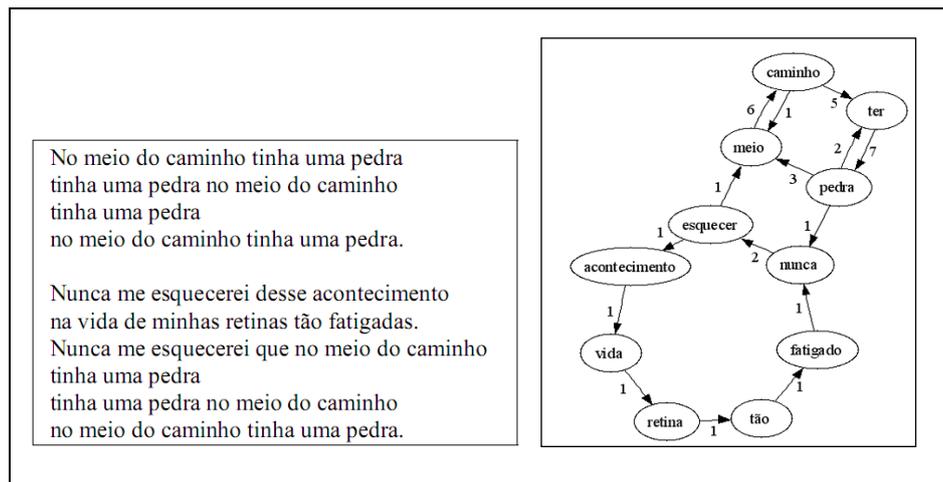


Figura 34 – Rede de ligação entre palavras adjacentes componentes de poema de Carlos Drummond de Andrade conforme análise de Antiqueira et al. (2005).

Os autores relatam três experimentos em que foram analisados diversos textos dos cadernos Dinheiro e Esporte do jornal Folha de São Paulo, redações da FUVEST e textos informativos produzidos por alunos do curso de letras da Universidade Mackenzie, SP. Processados os textos, efetuaram-se algumas medidas estatísticas a serem confrontadas com as notas atribuídas por um corpo de jurados a quatro critérios em cada texto: legibilidade, clareza, complexidade e qualidade. A conclusão a que se chegou é que, embora para textos julgados de boa qualidade as medidas estatísticas não representem indícios significativos, para os textos de pior valoração a “*qualidade diminui quando os caminhos mínimos aumentam, o que elucida a dificuldade dos escritores inexperientes em estabelecer conexões entre numerosos conceitos*”.

Objetivando o uso de redes conceituais na qualificação de práticas pedagógicas, Macedo (2010) relata uma técnica de mineração de texto em que após a etapa da eliminação das palavras de ligação, pode-se escolher a quantidade máxima de nós a serem considerados, bem como a frequência mínima de ocorrência de um termo para sua validação como um dos nós da rede: para um limite de 50 termos de frequência mínima igual a três, a rede será montada com os 50 termos mais frequentes ou algo menor que isso em caso de que dentre

estes termos mais utilizados constem alguns que tenham ocorrido apenas uma ou duas vezes ao longo do texto. Outra funcionalidade permitida pela técnica empregada é a possibilidade do trabalho com termos compostos por mais de uma palavra.

Com todos estes recursos em mãos, podem-se criar diversas versões de enredamentos representativos de um texto a fim de melhor investigar sua qualidade. Se, por exemplo, a construção da rede a partir dos 50 termos mais frequentes resultar em um mapeamento por demais fragmentado, pode-se ampliar para 75 o número máximo de nós a fim de verificar a existência de palavras que embora menos frequentes tenham alto impacto na conexão dos fragmentos.

Tendo aplicado todo esse processo a textos construídos coletivamente por alunos de cursos de extensão, graduação e pós-graduação, Macedo (2010) aponta sua eficiência na identificação da centralidade do texto para com o tema proposto. A imagem da rede permite, enfim, entender do que trata o texto, mesmo sem tê-lo lido ou, pelo contrário, inferir a perda de foco da produção textual nos casos em que a rede não indique um tema central. Outra percepção relatada pela autora é a correspondência entre a estrutura de redes e textos. Redes fragmentadas correspondem a textos também fragmentados, indicando a necessidade de uma ação pedagógica que possibilite o aprimoramento dos textos. Redes em que as conexões prevalecem indicam boa fluidez textual e a necessidade de uma prática pedagógica que aposte em maior autonomia do educando. As figuras 35 e 36 ilustram redes representativas de cada uma destas situações.

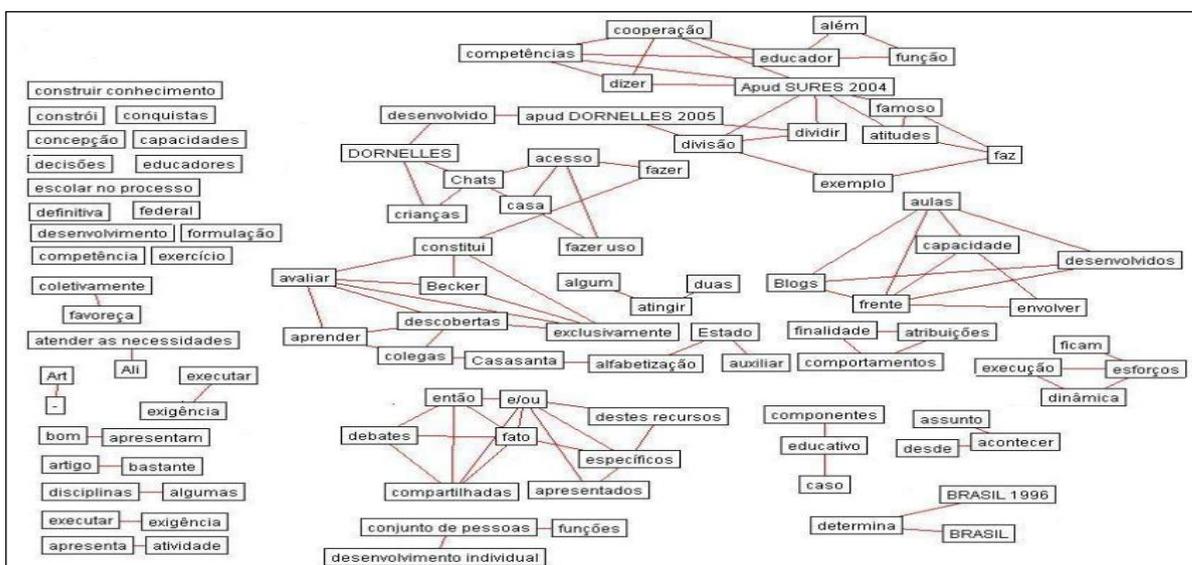


Figura 35 – Representação das conexões entre os 100 termos mais frequentes (com um mínimo de 2 ocorrências) componentes de um texto analisado por Macedo (2010) que, pela acentuada fragmentação da rede, apresenta falta de coesão, indicando a necessidade de uma ação pedagógica que possibilite seu aprimoramento.



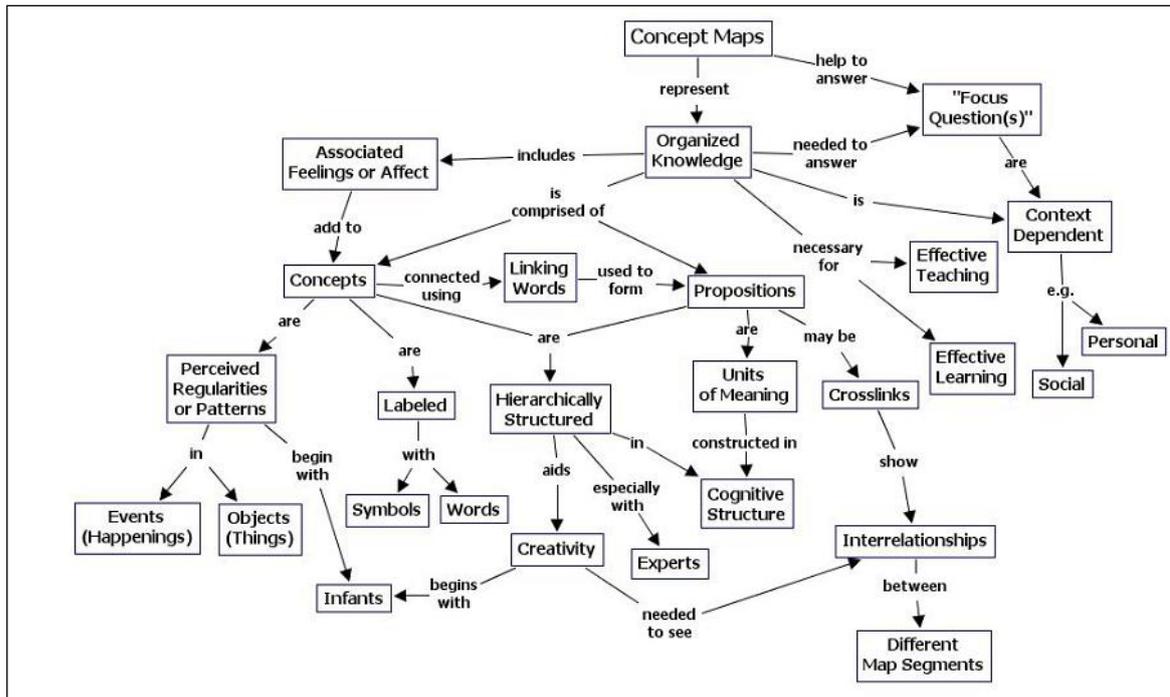


Figura 37 – Mapa conceitual relativo ao conceito, estrutura e objetivo de um mapa conceitual (CAÑAS et al, 2003).

### 3.9 O ENREDO COMO REDE

Enredos também podem ser visualizados como rede. É o que fez Moreti (2011) em sua investigação de *Hamlet*, interpretada como uma rede com nós atribuídos a seus personagens e *links* às interações explicitadas por meio dos diálogos existentes entre tais personagens ao longo de toda a peça. Segundo o autor, são vários os ganhos que a visualização dessa rede permite, com a espacialização do tempo sendo uma delas: quando se assiste ou se lê a peça viaja-se pela sequência cronológica dos acontecimentos; por sua vez, na visão em rede todos os personagens aparecem juntos, com o passado ficando tão visível quanto o presente, permitindo uma visão global dos espaços de cada personagem na tragédia. Segundo Moreti (2011), como consequência dessa espacialização, visualizam-se subsistemas como a região de todas as tramas de assassinatos presentes na peça (ver figura 38):

Fazendo o passado tão visível como o presente, esta é uma grande mudança introduzida pela utilização de redes. Assim, elas fazem visíveis "regiões" específicas dentro do enredo como um todo: subsistemas que compartilham alguma propriedade significativa. Considere os personagens ligados tanto a Cláudio [Claudius, no original] como a Hamlet na figura [38]: exceto para Osric e Horácio [Horatio, no original], cujo link para Cláudio é, todavia, extremamente tênue, eles estão todos mortos. Morto por quem, nem sempre é fácil dizer: Polônio [Polonius, no original] é morto por Hamlet, por exemplo - mas Hamlet não tem ideia de que é Polônio quem ele está apunhalando atrás da cortina, Gertrudes [rainha e mãe de Hamlet] é morta

por Cláudio [rei e padraсто do príncipe Hamlet] - mas com veneno preparado para Hamlet, não para ela; Hamlet é morto por Laertes, com a ajuda de Cláudio, enquanto o próprio Laertes, tal qual Rosencrantz e Guildenstern antes dele, são todos mortos por Hamlet, mas com as armas de Cláudio. A ação individual é confusa; o que é verdadeiramente mortal é a posição dos personagens na rede, acorrentados aos polos rivais do rei e do príncipe. Fora da região vermelha, ninguém morre em Hamlet. A tragédia está toda lá. (MORETI, 2011, tradução nossa).

Moreti (2011) destaca que a visão em rede permite uma percepção não usual do protagonista, que passa a ser o personagem que minimiza a soma das distâncias para todos os outros vértices. Na figura 38 se percebe que Hamlet está a um grau de separação de 16 personagens e a dois graus de todos os outros.

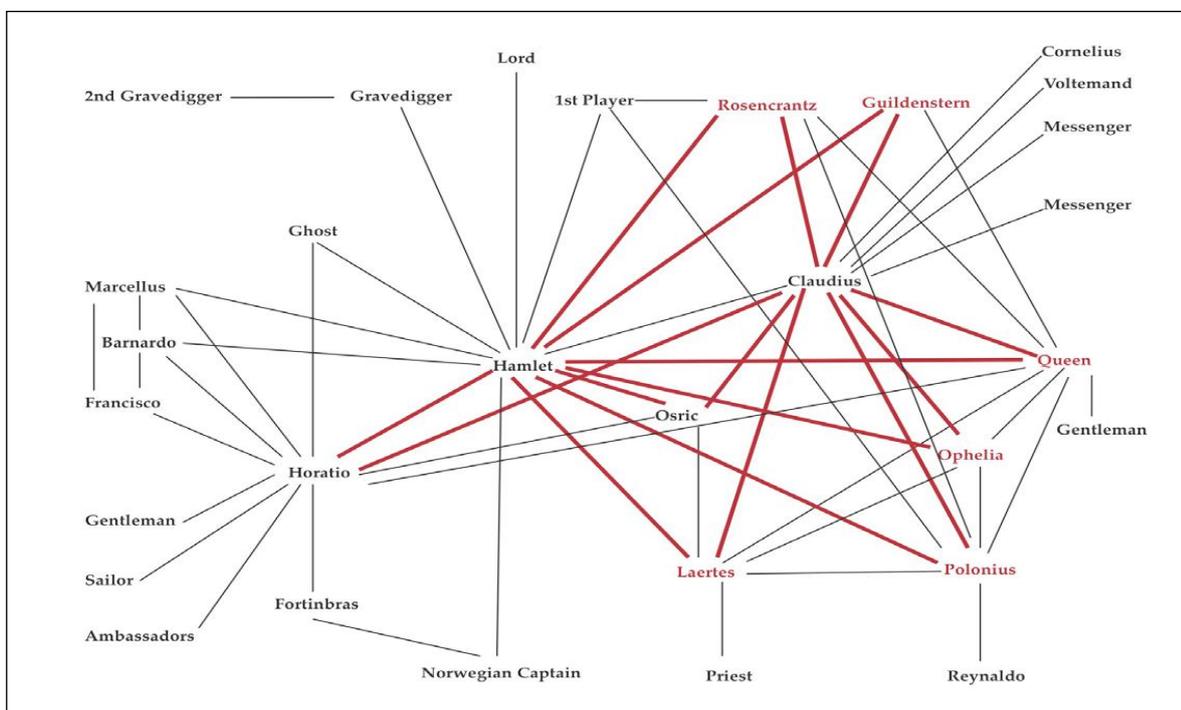


Figura 38 - “Espaço das mortes” em *Hamlet*

Além disso, um histograma (figura 39) indicando a quantidade de *links* associado a cada personagem, revela a lei de potência característica de muitas redes: poucos nós com muitos *links* e muitos nós com poucos *links*. Nesse sentido, o autor argumenta quanto à incoerência de se procurar definir um típico personagem no enredo: “então, falar dos personagens de Shakespeare “em geral” é errado, ao menos nas tragédias, porque estes “personagens em geral” não existem: tudo o que há é uma curva [no histograma] levando de um extremo a outro sem qualquer evidente solução de continuidade” (MORETI, 2011, tradução nossa).

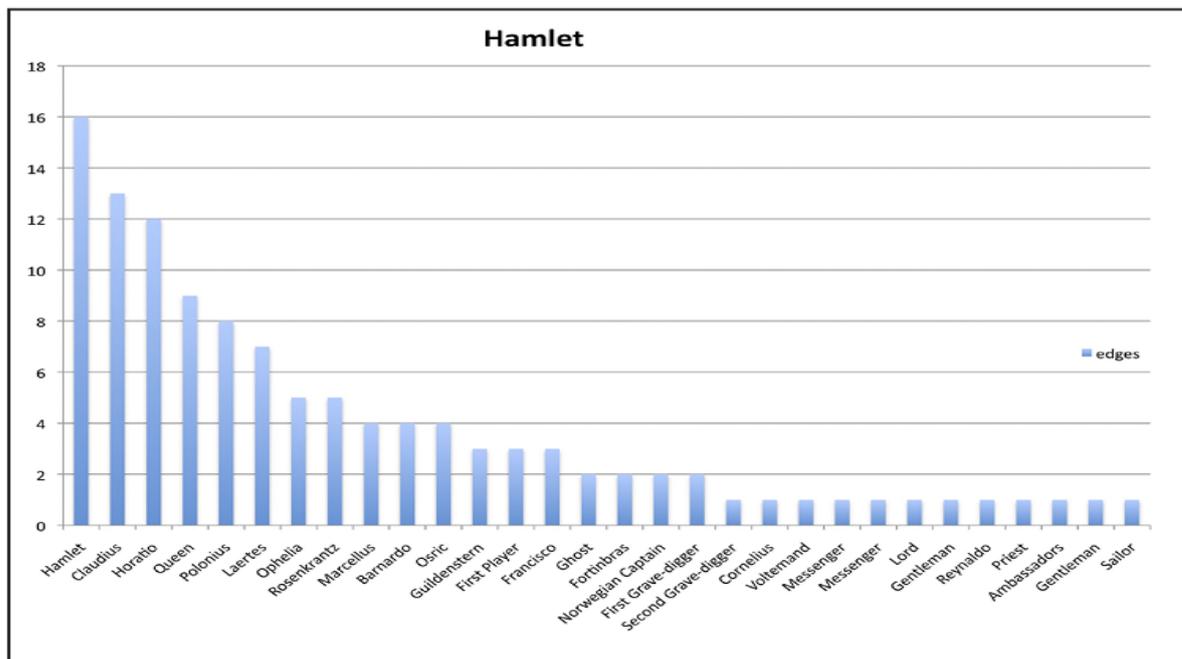


Figura 39 - Centralidade de cada personagem em *Hamlet*; na vertical vê-se o número de conexões de cada personagem ao longo do enredo.

Moreti (2011) observa também que a diferença entre os graus de cada personagem pode ser realçada por experimentos de intervenção no enredo, como se ele fosse um modelo. É possível, por exemplo, pensar em *Hamlet* sem Hamlet. Tal experimento é ilustrado na figura 40, resultando na nítida fragmentação da rede em dois setores ligados por apenas três links.

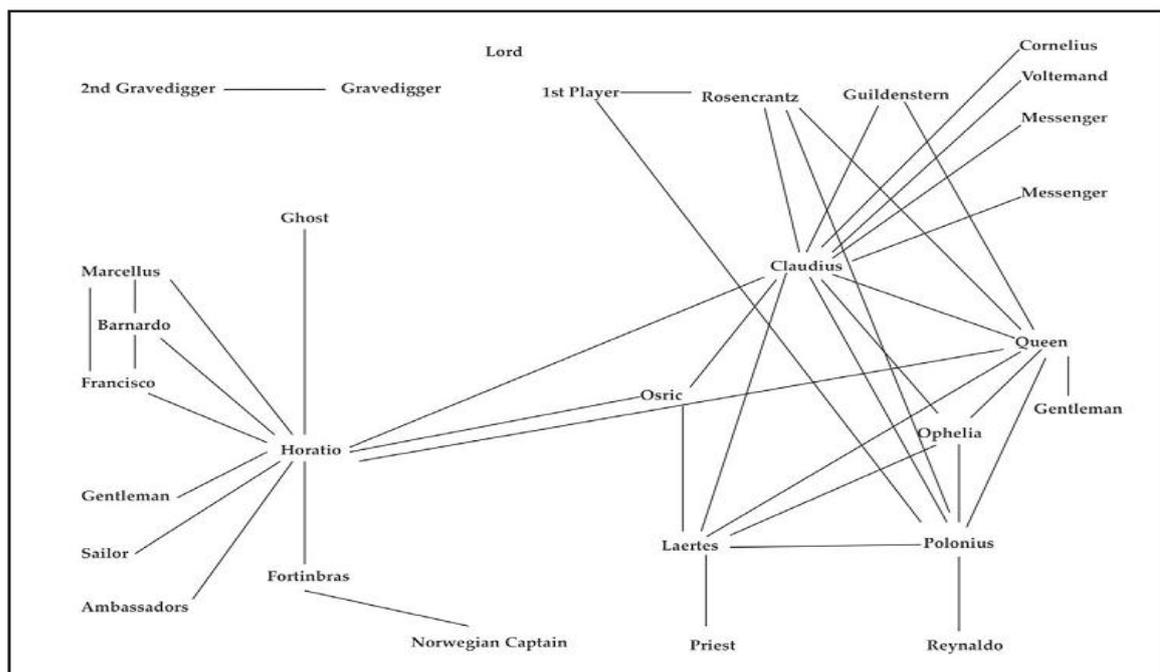


Figura 40 - O “Espaço das mortes” em *Hamlet* sem Hamlet.

No entanto, quando o mesmo experimento é feito com Cláudio, tal qual ilustrado na figura 41, a fragmentação resultante é muito pequena, em que pese ser Cláudio o personagem de maior grau após Hamlet.

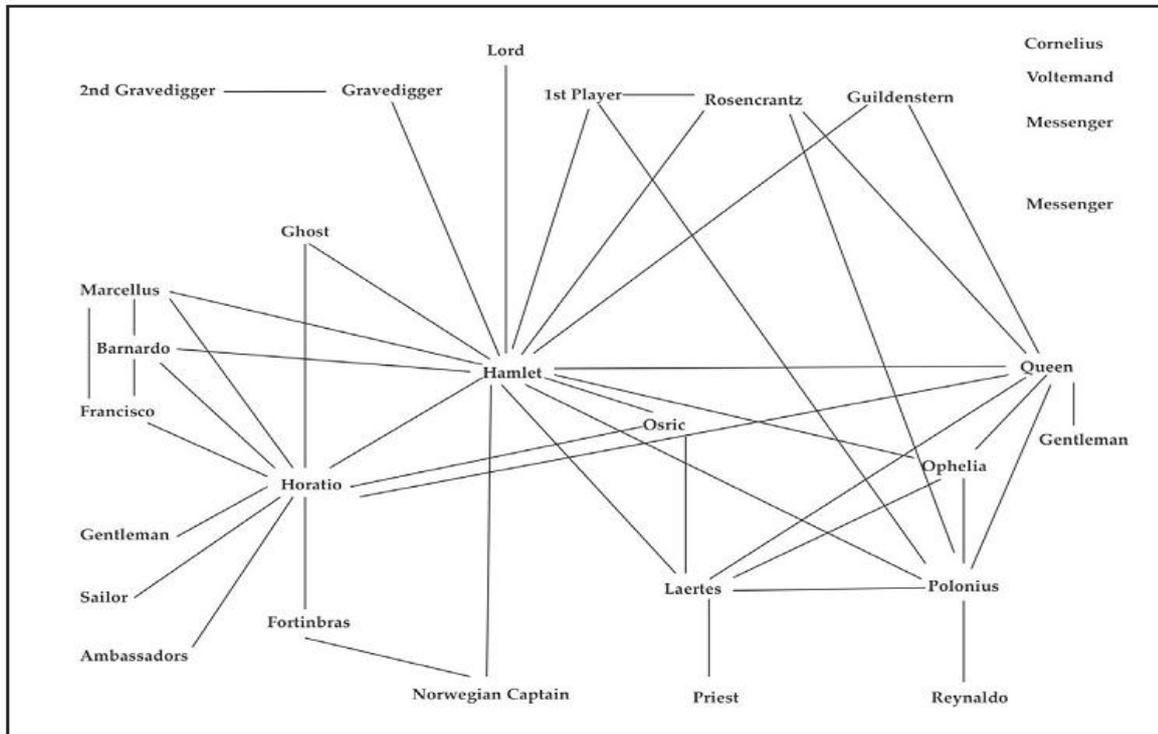


Figura 41 - O “Espaço das mortes” em *Hamlet* sem Cláudio.

A explicação reside na localização dos dois personagens na rede: enquanto Cláudio está dentro da corte, com cerca de 50% das palavras por ele dita ao longo da peça restringindo-se ao sexteto formado em conjunto com a Rainha, Laertes, Ofélia, Polônio e Hamlet (figura 42), este último compartilha com tais personagens apenas 28% de sua fala, tendo papel ativo tanto dentro como fora da corte. Em termos técnicos, tal sexteto é uma região de alta aglomeração ou *cluster*, com todos os personagens relacionando-se diretamente entre si e assim permanecendo mesmo após a retirada de Cláudio da rede. Hamlet, por sua vez, ocupa papel proeminente na estrutura da rede como um todo.

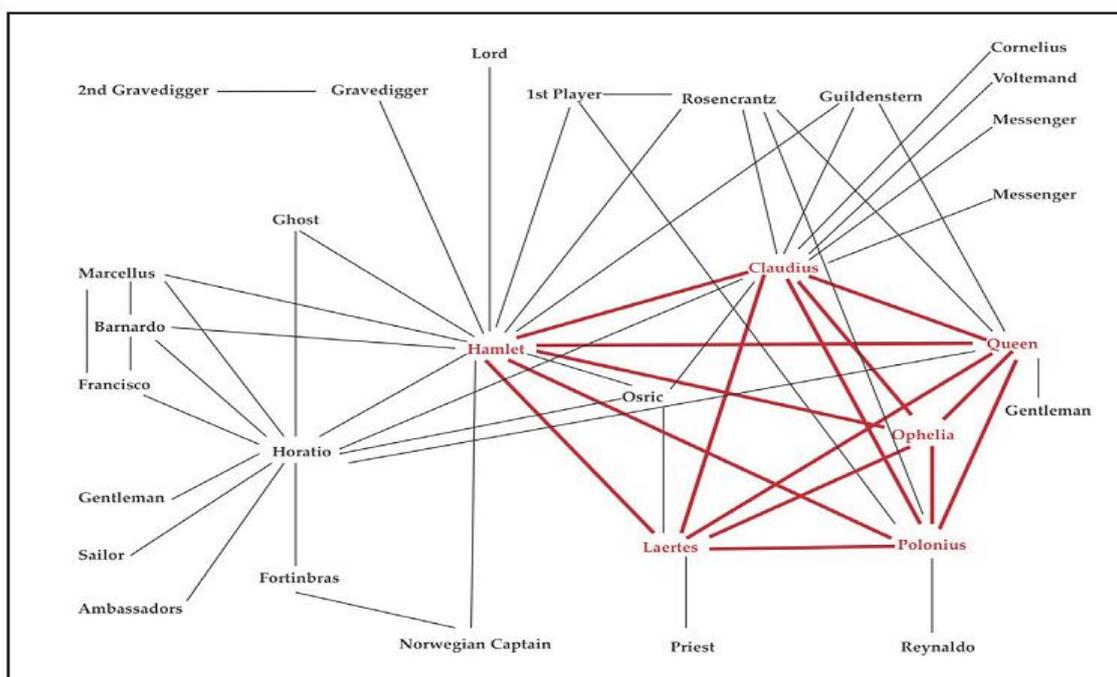


Figura 42 - O sexteto interno à corte em *Hamlet*.

### 3.10 HISTÓRIAS CONTRAFACUAIS E REDES DE INFLUÊNCIAS ENTRE AVANÇOS

Embora sem neles constar menção explícita às pesquisas sobre as redes complexas, os trabalhos desenvolvidos por Pessoa Jr acerca dos mapeamentos de histórias contrafactuais podem também ser resgatados neste capítulo. Afinal, similar ao mapeamento de coocorrências de citações, a pesquisa por histórias contrafactuais envolve a análise da correlação de assuntos e citações presentes em centenas de artigos de forma a construir imagens como a figura 5, apresentada no item 2.1 (página 29). Esclarece Pessoa Jr (2000), no entanto, que o aspecto fundamental do método é o mapeamento da rede de influências entre avanços, isto é, entre as unidades passadas de cientista para cientista:

[ ] Um “avanço” não é necessariamente um passo positivo, na direção correta, conforme sugerido pelo significado usual do termo. Para nós, os avanços são as unidades que são passadas de cientista para cientista, são os elementos que são adicionados ao conjunto de ideias, dados, leis, informações, conhecimentos tácitos etc., disponíveis para um certo cientista em uma época específica. Cada cientista assimila um conjunto de avanços, seleciona alguns, rejeita temporariamente outros, combina dois ou mais avanços etc. Segundo nossa visão, a ciência evolui a partir desses avanços disponíveis e dos novos avanços imaginados ou descobertos pelos cientistas teóricos e experimentais. Tal concepção é próxima a de algumas correntes da chamada Epistemologia Evolucionária [...].

No entanto, salientamos que os avanços são também as unidades que se conservariam na passagem de uma história possível para outra. Após determinar um conjunto de avanços concatenados em uma rede de influência para a história factual,

passamos a postular histórias contrafactuais imaginando ordenamentos diferentes para os mesmos avanços. Naturalmente, novos avanços também deverão ser postulados para diferentes histórias. (PESSOA JR, 2000).

Não se trata, portanto, de um “simples” mapeamento das coocorrências de citações, mas na “dissecação” de cada texto de forma a identificar o tipo de artigo (se experimental ou teórico, resenha, carta, coleta de dados, etc), seus antecedentes (trabalhos que o influenciaram diretamente) e os avanços nele obtidos (cálculo de constante, ampliação de domínio, anomalia teórica, distinção teórica, lei empírica, método experimental, etc). De posse de todos esses dados, desenha-se uma rede com os avanços como nós conectados por suas influências diretas, sem referência a artigos ou cientistas. Nesta rede de influências, esclarece Pessoa Jr (2000), os avanços “*devem ser pensados como um conjunto de dados disponíveis na literatura (independente de quem os obteve), dados estes que são concatenados (pelos cientistas) com outros avanços, tendo em vista a geração de novos avanços*”.

A figura 43 ilustra a rede de influências entre avanços para o caso da radiação térmica. As linhas tracejadas que nela aparecem indicam uma relação de influência fraca entre dois avanços, correspondente às situações em que o avanço decorrente poderia ter ocorrido de forma independente do avanço antecedente. Pelo contrário, linhas cheias indicam uma relação de influência forte, com a identificação de que o avanço antecedente foi necessário para a ocorrência de seu decorrente. Por sua vez, linhas que intercalam traços e pontos (-·-·-·-) indicam a conexão entre uma lei empírica e dados experimentais que confirmam ou negam esta lei, inferindo a plausibilidade de histórias contrafactuais em que os dados experimentais pudessem ter sido obtidos sem a existência prévia da lei empírica.

Segundo Pessoa Jr (2000), é o levantamento de redes de influências entre avanços que possibilita aplicar quatro estratégias intuitivas simples para visualizar diferentes histórias possíveis: (a) suposição de que um avanço aparentemente “improvável” não tivesse ocorrido; (b) suposição de que um avanço retrospectivamente “provável” tivesse ocorrido antes do tempo; (c) identificação de descobertas independentes, isto é, avanços iguais ou semelhantes que se deram de maneira independente um do outro; (d) suposição de que a consequência de uma descoberta fundamental (na história factual) tenha sido (numa história contrafactual) a causa desta descoberta.

Além da radiação térmica, o autor relata que outras três áreas foram alvo de sua pesquisa e consequente construção das respectivas redes de influências entre avanços: espectroscopia, efeitos ópticos e calores específicos. São justamente estas quatro redes que



### 3.11 CONEXÕES

Ainda que existam muitas outras redes com investigações já publicadas<sup>17</sup>, o recorte apresentado nesse capítulo fornece boa percepção das produções de um campo de pesquisa que tem se construído por interações com e entre diversas áreas científicas. Nesse sentido, vale realçar conexão com a exposição realizada no capítulo anterior: a quântica não só é base tecnológica da microeletrônica, como também de diversos equipamentos utilizados e fenômenos estudados por algumas das pesquisas de mapeamento de redes, haja vista as técnicas utilizadas na medicina para obtenção de imagem de atividades cerebrais.

Mas a discussão aqui apresentada, especialmente no que se refere às técnicas de mineração de texto e construção de redes de conceitos, também se conecta, em combinação com a exposição do próximo capítulo acerca da concepção do conhecimento como rede de significados, com algumas das práticas na formação de professores apresentadas no Capítulo 5. Em síntese, a ideia que será vista na formação de professores fará uso explícito daquelas técnicas na obtenção de uma visão global dos textos e redes conceituais construídos pelo conjunto de licenciandos participantes de uma disciplina focada na física quântica enquanto conhecimento em rede, bem como na análise do texto *Radiações, materiais, átomos e núcleos* expressivamente utilizado ao longo de um curso sobre estrutura da matéria.

Conforme citado na introdução, a percepção de tal conexão não estava dada, no entanto, no início deste doutoramento, com a aquisição da visão geral sobre a teoria de redes e suas aplicações tendo sido essencial para sua obtenção. Esse é, aliás, o motivo pelo qual se preferiu apresentar a visão panorâmica de todo um conjunto de pesquisas do que apostar em caminho mais direto rumo ao componente efetivamente aplicado. Em suma, por homologia, procede-se em rede para explicitar as redes...

---

<sup>17</sup> Algumas delas: relacionamentos pessoais (como o sexual), esportivas (como a de jogadores de futebol conectados por atuação simultânea em um mesmo clube), musicais (como a de compositores brasileiros conectados por terem suas canções interpretadas por um mesmo cantor), transportes (como a de aeroportos conectados pelos voos entre eles), eletrônicas (como as conexões entre chips, transistores, resistores e outros dispositivos pertencentes a um circuito eletrônico), fenômenos naturais (como a de duas regiões da superfície terrestre conectadas pela ação subsequente de terremotos), matemáticas (como a de números naturais conectados por ao menos um fator primo em comum) e epidêmicas (sejam elas espalhadas por vírus biológicos ou computacionais).