

Redes Complexas - Motifs em Redes Complexas

Gustavo Vrech, Luiz Gonzaga

Instituto de Física de São Carlos, USP, 13566-590, Brazil

14 de Junho de 2011

1 Introdução

Uma rede complexa é uma estrutura matemática que baseia-se num conjunto de elementos com características comuns (vértice) e uma regra de interação entre esses elementos (aresta). Deste modo, praticamente qualquer sistema poderia ser descrito através de uma rede complexa, justificando a grande variedade de aplicações em inúmeros campos da ciência.

Conforme discutido nos tópicos iniciais de [1], pode-se utilizar as redes complexas para modelar redes sociais, como relações de amizades como no experimento de Milgram, redes de relações comerciais, casamentos entre famílias, e até mesmo redes de contato sexual. Redes de informações também podem ser estudadas, como por exemplo, a rede formada pelas citações entre artigos científicos, ou a World Wide Web, a rede de páginas de internet conectadas entre si a partir de hiperligações. A estrutura física da rede de computadores, seu cabeamento e roteadores, também pode ser interpretados como uma rede complexa, assim como a estrutura da rede de distribuição elétrica de uma região, ou mesmo a conexão dos vôos entre aeroportos.

Já um motif é um pequeno elemento de um todo, que pode ser analisado para dar informações sobre todo um conjunto. Em música, um motif, ou motivo é um pequeno trecho de melodia que permite a um ouvinte identificar qual música está sendo reproduzida escutando apenas estas poucas notas. No caso das redes complexas, um motif é uma sub-rede que de algum modo é representativa para a rede, um pequeno elemento básico como um tijolo, que de algum modo caracteriza toda a construção.

Um motif é definido como uma sub-rede de uma rede maior, um subconjunto de qualquer tamanho. Tipicamente procuram-se os motifs de pequeno tamanho, aproximadamente de três a cinco nós ou arestas. Qualquer rede poderia ser subdividida em todos os seus motifs, e assim feita o levantamento estatístico sobre sua composição tal qual um espectro de frequências, porém as redes que possuem motifs mais significativos, aqueles que aparecem muito mais do que os outros evidenciam o real potencial prático dos motifs.

Diversos artigos têm recentemente tratado dos motifs, em especial ao analisar redes biológicas. O primeiro artigo notável sobre a área [2] retrata um panorama sobre a área, oferecendo especial

importância para diversos tipos de aplicações. No presente trabalho é feita uma revisão sobre diversos trabalhos que lidam com os motifs, inicialmente identificando motifs básicos presentes em diversas redes. Em seguida é apresentado um algoritmo padrão utilizado para a detecção de motifs, assim como algumas de suas características e especificidades.

Em seguida são resumidos três trabalhos de aplicações de motifs, sendo que estes foram escolhidos devido às suas diferenças de aplicações, assim como de procedimento tratado. O primeiro trabalho aborda aspectos biológicos dos motifs, como a interação gênica pode ser modelado a partir deste conceito. O segundo lida com a rede de informação da Wikipédia, mais especificamente da topologia formada pelas relações de pagina-autoria, assim como os motifs que diferenciam dois tipos de redes, as dos jogadores de futebol e sociólogos. O ultimo trabalho mostra como uma rede formada a partir de uma música típica japonesa pode ser decomposta em motifs gerando músicas aleatórias de sonoridade agradável.

2 Exemplos de Motifs

Alguns tipos de motifs ocorrem tão freqüentemente em diversos sistemas e em especial em sistemas biológicos, que merecem uma atenção e discussão especial fora do seu contexto de aplicação direta. Nessa seção serão descritos os motifs denominados *auto-regulation*, *feedforward loop*, *bi-fan* e *bi-parallel*. A figura 1 exemplifica cada um dos motifs aqui apresentados.

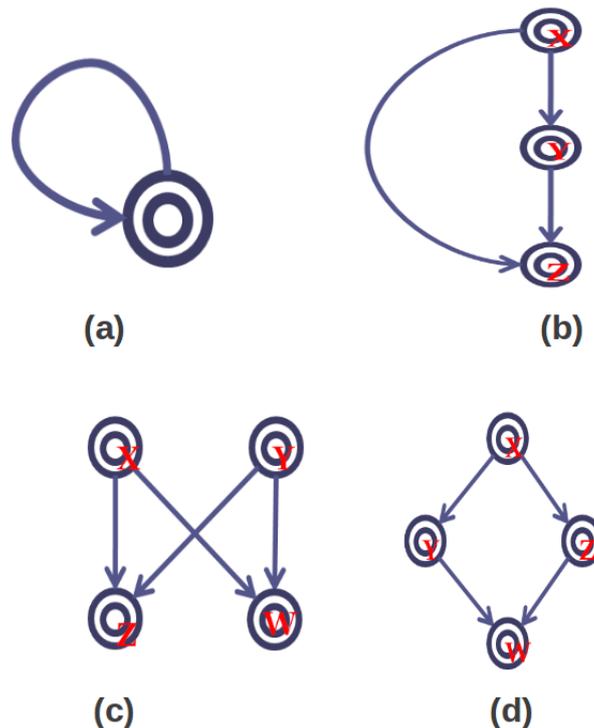


Figura 1: Exemplos de motifs: (a) auto-regulation, (b) feedforward loop, (c) bi-fan, (d) bi-parallel

2.1 Auto-regulation motif

Este motif consiste num único nó com uma ligação para si próprio. Ele aparece freqüentemente nas redes da bactéria *E. coli*, e seu papel é a regulação da sua própria transcrição [3]. Sua ligação

pode assumir dois valores, um positivo, tornando assim a auto-regulação positiva, aumentando sua taxa de transcrição, tanto negativa, diminuindo esta taxa. Sua representação esquemática pode ser vista na figura 1(a).

2.2 Feedforward loop modif

Provavelmente o motif mais comum em redes biológicas, este consiste em três aqui denominados como X, Y e Z. O nó Z é influenciado pelos nós X e Y, enquanto que o nó Y é influenciado unicamente pelo nó X, conforme a figura 1(b). Este motif é encontrado em redes genéticas, redes baseadas em cadeias alimentares, redes de sinapses neuronais, redes de circuitos lógicos, etc, qualquer rede onde exista uma cadeia de interações e onde um dos elementos dessa cadeia pode influenciar mais do que apenas um nó.

De forma semelhante a apresentada na seção anterior, as arestas deste motif podem ter um efeito regulatório positivo ou negativo, resultando em oito formas distintas do motif. Pode-se dividir dois grupos de motifs feedforward loop, os coherentes e os incoherentes. O coerente é aquele em que todas as arestas são positivas, enquanto que os incoherentes têm uma ou mais interações negativas, podendo proporcionar ou não equilíbrio entre os componentes. No caso dos circuitos lógicos, o tipo coerente é utilizado para o gate lógico AND.

2.3 Bi-fan modif

O motif Bi-fan é apresentado na figura 1(c), consiste em dois nós reguladores, X e Y e dois nós regulados, Z e W. Existem arestas ligando todos os nós reguladores aos regulados. Este motif é comum em redes onde uma determinada classe de objetos (reguladores) atua diretamente sobre uma outra classe (regulados), num regime onde qualquer regulador pode regular qualquer nó regulado de maneira indiscriminada.

De modo bastante semelhante ao feedforward loop, este motif é dividido em dois subgrupos, os coherentes e incoherentes. No caso do Bi-fan em específico, existe a possibilidade de um motif ser parcialmente coerente, quando apenas um dos nós regulados recebe arestas negativas. Este motif ocorre em Redes genéticas, protéicas, neuronais, e em circuitos de lógica.

2.4 Bi-parallel motif

Este tipo de motif é bastante comum em redes que representam cadeias alimentares, quando um mesmo predador se alimenta de dois tipos diferentes de preza, porém ambas as presas tem o mesmo hábito alimentar. Ela também ocorre em circuitos lógicos e redes neuronais. Sua topologia é evidenciada na figura 1(d).

3 Algoritmos

3.1 Concentração de Subgrafos

Considerando redes dirigidas e apenas um tipo de aresta e nó, define-se o número de aparição de subgrafos (N_i) do tipo i [4]. A concentração de subgrafos n-nó do tipo i é a razão entre o número das aparências e do número total de n-nó subgrafos conectados em rede:

$$C_i = \frac{N_i}{\sum_i N_i} \quad (1)$$

3.2 Amostragem de Subgrafos

O algoritmo de amostragem de subgrafos trabalha escolhendo aleatoriamente arestas conectadas até que um conjunto de n nós seja atingido. A seguir descreveremos o procedimento de amostragem aleatória de um subgrafo n -nó da rede: seleciona uma aresta aleatória da rede e, em seguida, expande o subgrafo iterativamente escolhendo aleatoriamente arestas vizinhas até o subgrafo atingir n nós. Para cada escolha aleatória de uma aresta, a fim de pegar uma aresta que irá expandir o tamanho do subgrafo por um lado, prepara-se uma lista de todas as arestas candidatas, e depois de escolher aleatoriamente uma aresta da lista. Finalmente, o subgrafo “amostra” é definido pelo conjunto de n nós e todas as arestas que se conectam entre esses nós da rede original (e não apenas as arestas que foram escolhidos pelo processo de expansão). (Veja a descrição formal de algoritmo na Figura 3.2)

Definitions: E_S is the set of picked edges.
 V_S is the set of all nodes that are touched by the edges in E_S .

Init V_S and E_S to be empty sets.

1. Pick a random edge $e_1 = (v_i, v_j)$. Update $E_S = \{e_1\}, V_S = \{v_i, v_j\}$
2. Make a list L of all neighboring edges of E_S .
 Omit from L all edges between members of V_S . If L is empty return to 1.
3. Pick a random edge $e = (v_k, v_l)$ from L .
 Update $E_S = E_S \cup \{e\}, V_S = V_S \cup \{v_k, v_l\}$
4. Repeat steps 2–3 until completing n -node subgraph S .
5. Calculate the probability P to sample S .

Figura 2: Algoritmo amostragem de subgrafo

3.3 Correção exata de amostragem não-uniformes

Um subgrafo específico é um conjunto de n nós conectados na rede. As probabilidades de amostragens de diferentes subgrafos específicos na rede não são iguais, mesmo que tenham a mesma topologia. A fim de corrigir isto, podemos calcular a probabilidade, P , de um subgrafo de amostragem específica. Cada tipo subgrafo recebe uma pontuação. Depois de cada amostragem, adiciona-se uma pontuação ponderada de $W = 1/P$ para a pontuação do tipo subgrafo relevantes. Este processo é repetido para um número total de amostras de ST . Finalmente, podemos calcular as concentrações de todos os tipos de subgrafos de acordo com suas pontuações.

Para uma amostra de subgrafos n -nós, um conjunto ordenado de $n - 1$ arestas é iterativamente escolhidos aleatoriamente. Para calcular a probabilidade, P , por amostragem no subgrafo, precisamos verificar todos esses possíveis conjuntos ordenados de $n - 1$ arestas [denotado como $(n - 1)$ -permutações] que poderia levar a amostragem do subgrafo. A probabilidade de amostragem do subgrafo é a soma das probabilidades de todos esses possíveis conjuntos ordenados de $n - 1$ arestas:

$$P = \sum_{\sigma \in S_m} \prod_{E_j \in \sigma} Pr[E_j = e | (E_1, \dots, E_{j-i}) = (e_1, \dots, e_{j-1})] \quad (2)$$

Onde S_m é um conjunto de todos os $(n-1)$ -permutações das arestas, a partir da aresta de um subgrafo específico que, poderiam levar a uma amostra do subgrafo. E_j é a j -ésima aresta em uma específica $(n-1)$ -permutação (σ). Na Figura 3.3, que ilustra este procedimento em uma rede simples. Os dois subgrafos específicos considerados neste exemplo, têm diferentes probabilidades de amostragem e são atribuídos pesos diferentes a fim de garantir estimativa imparcial de concentrações subgrafo.

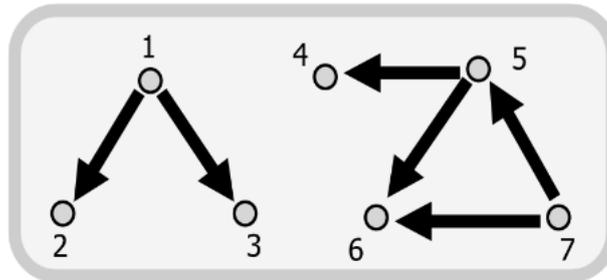


Figura 3: Subgrafos específicos

3.4 Calcular as concentrações de subgrafos n-nó

Define-se uma pontuação S_i para cada tipo de subgrafo i . Inicialmente esta pontuação é zero para todos os tipos de subgrafos. Para cada amostra é adicionado uma pontuação de ponderação $W = 1/P$ com a pontuação acumulada, S_i , do i tipo subgrafo relevante: $S_i = S_i + W$. Após as ST amostras, assumindo que nós evidenciamos L tipos diferentes de subgrafo, podemos calcular as concentrações estimadas do subgrafo:

$$C_i = \frac{S_i}{\sum_{k=1}^L S_k} \quad (3)$$

Obs: O algoritmo pode evidenciar o mesmo subgrafo múltiplas vezes.

3.5 Enumeração Randômica

Um novo algoritmo que é utilizado de forma eficiente enumera todos os subgrafos de tamanho k . Este algoritmo é modificado de forma a “pular” aleatoriamente alguns dos subgrafos durante a execução. Este novo algoritmo é comprovado ser superior ao algoritmo apresentado anteriormente.

4 Motifs em interações genéticas

Diferentes modos de interação genética indicam diferentes tipos de relações funcionais entre os genes [5]. A extração de informações biológicas a partir de redes de multi-modo de interação genética exige métodos adequados de estatística e informática. Motifs extraído a partir da rede de multi-modos de interação genética formam sub-redes funcionais, genes dominantes destacam-se nessas sub-redes, e revelam as reflexões genética do sistema bioquímico subjacente.

4.1 Background

A célula é uma elaborada rede de interações biomolecular e ambiental que trazem juntos fenótipos complexos. Entender as consequências funcionais das interações moleculares é fundamental para a

compreensão de fenótipos. Uma abordagem altamente bem-sucedido é o uso de interações genéticas. Interações genéticas descrevem as consequências fenotípicas de combinações de genes. Interações genéticas combinadas com dados de interação molecular pode delinear os fluxos de informação através de sistemas bioquímicos complexos. O conceito da via de sinalização molecular deve muito a esta abordagem.

A interação genética inclui medições fenotípicas de quatro genótipos: o genótipo de referência (wild type (WT)); uma única perturbação A de um gene, uma perturbação B de um gene diferente, e uma dupla perturbação AB. Por si só, as perturbações única, ligam genes individuais a características fenotípicas específicas e processos biológicos. Estudo de uma perturbação dupla define as relações funcionais entre os genes perturbado. A ordem relativa das quatro medições fenótipo define os diferentes modos de interação genética. Modos de interação genética indicam uma ou mais possíveis relações moleculares, por exemplo, a *upstream/downstream*. Redes de interação genética e emaranhado molecular, restringem essas possibilidades. Desta forma, modos de interação genético são um reflexo do sistema bioquímico subjacente.

Os geneticistas têm formalizado coleções de interações genéticas em redes de interação, onde os nós representam os genes perturbados e as arestas as interação genética. Neste estudo, os padrões de rede foram utilizados para reduzir o sistema global num mapa temático de relações biológicas. O método E-MAP [5] cria redes de alta densidade de interação genética consistindo em agravar ou atenuar os tipos de arestas.

Mais trabalho gerou redes de múltiplos modos de interação entre o genótipo (tipos de aresta). Em Drees et al. [6], todas as possíveis interações genéticas foram classificadas em nove modalidades, dos quais quatro são assimétricas (arestas). Uma rede de multi-modos de intereção genética foi derivada de um grande conjunto de dados fenótipos quantitativos. Revelando padrões de interações genéticas local e global sugerindo a relevância da informação contida na estrutura e distribuição de interações genéticas dentro da rede. Mais informações sobre a rede pode ser extraído de tais redes complexas, identificando de forma significativa os padrões de interação genética, *network motifs*.

4.2 Rede de multi-modo de interação genética

Na rede utilizada, existem 1.760 interações genéticas entre os 128 genes que controlam o fenótipo de invasão de agár no brotamento da levedura. As perturbações incluem deleção do gene, assim como os alelos dominantes mais expressivos. Esta rede contém todas as nove modalidades possíveis de interação genética, epistáticos, sintético, supressivas, aditivas, condicional, não-sintética, não monotônica, a dupla interação não monotônica. Quatro desses modos (epistáticos, supressivas, condicional e não monotônica) são direcionais, dando treze possíveis arestas entre qualquer par de nó.

4.3 Padrões de interação genética refletem o subjacente sistema molecular

Antes de uma análise estatística rigorosa dos motifs, foram vistoriadas as redes de invasão de leveduras para distinguir padrões possíveis de interações genéticas refletindo o sistema de base molecular. A Figura 4 mostra as interações genéticas entre os componentes das três principais vias de sinalização de controlo de invasão de leveduras. Posteriormente, foi investigado as observações preliminares (descritas abaixo) quantitativamente e no mundo na rede.

Inicialmente observou-se que existem padrões locais incorporando tanto o tipo de aresta quanto a topologia da rede. Por exemplo, considerando as interações entre os genes mais expressivos, Cdc42 e GLN3 e as exclusões de DIG2 e TPK2. Ambos Cdc42 e GLN3 interagindo não-sinteticamente com DIG2 e não-monotonicamente com TPK2, criando padrões de dois modos de interação bifan. Além disso, observamos que os padrões de interação genética pode refletir o sentido do fluxo de

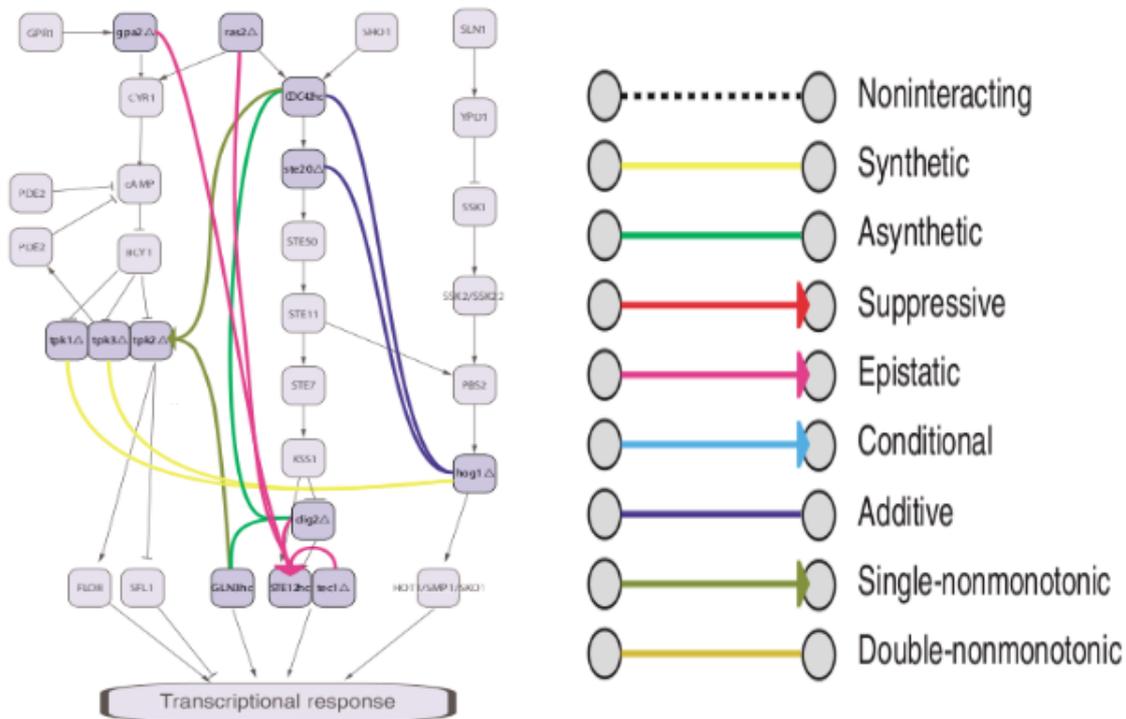


Figura 4: principais vias de sinalização de controle de invasão de leveduras

informações através da rede molecular.

4.4 Modelo estatístico

Padrões de interação genética biologicamente relevantes podem ser identificados encontrando aqueles que ocorrem com mais frequência na rede genética do que o esperado ao acaso. Isso pode ser feito comparando o número de vezes que um determinado padrão ocorre na rede genética para o número de vezes que ele ocorre em um conjunto de redes adequadamente randomizados. Desta forma, a significância pode ser atribuída a cada padrão identificado. Neste estudo destacam-se os padrões com um nível de significância de $(p \text{ menor } 0.05/n)$, onde n é o número de padrões testados em cada análise. Algoritmos foram desenvolvidos para criar um conjunto de redes randomizadas modelando a hipótese nula. A rede de invasão da levedura contém nove tipos de arestas, dos quais quatro são dirigidas. Redes randomizadas foram geradas por um método de *Monte Carlo* iterativamente selecionando um par de arestas de forma aleatória e trocando seus tipos de arestas.

Aleatorizações foram sujeitas a restrições específicas para impedir a introdução de obliquidades nos resultados. Cada aresta representa os resultados de um determinado experimento (medição repetida de fenótipos de WT, A, B e AB).

4.5 Motifs na rede de interação genética

Para identificar padrões de interação genética na rede, que reflete as relações biológicas, tais como as ilustradas na Figura 5, foram identificados *network-motifs*. Como já foi dito *network-motifs* são pequenas repetições que ocorrem na rede, onde a repetição sugere um significado funcional. Os padrões mais simples de rede, que contém informação sobre os modos de interação genética e sua organização em nível de sistema, são motifs de 3 nós (*3n-motifs*).

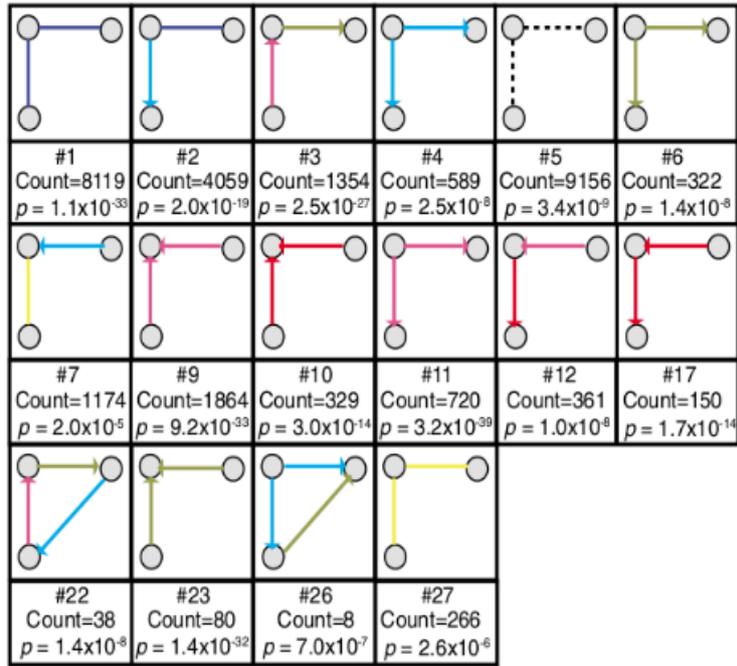


Figura 5: motifs de 3 nós (3n-motifs)

Motifs do tipo de aresta homogênea foram frequente, com 9 dos 13 possíveis padrões homogêneos sendo significativo (3n-motifs 1, 4, 5, 6, 9, 10, 11, 23, 27). Exemplos de tais motivos ocorrem na Figura 4. Muitos motifs heterogêneos também foram encontrados (3n-motifs 2, 3, 7, 8, 12 e assim por diante), assim como vários motifs totalmente conectado (por exemplo, 3n-motivos 22, 24, 25, 26).

Tambem foram identificados padrões de 4 nós (4n-motifs). Assim, um algoritmo de amostragem (materiais e métodos) foi empregado. Das 1.505 amostras de padrões 4n da rede original, 190 (12,6%) foram repetidos de forma significativa. Figura 6 mostra exemplos. Encontramos *4n-motifs* exibindo o tipo de aresta homogênea detectada entre 3n-motivos, bem como motifs de aresta mista.

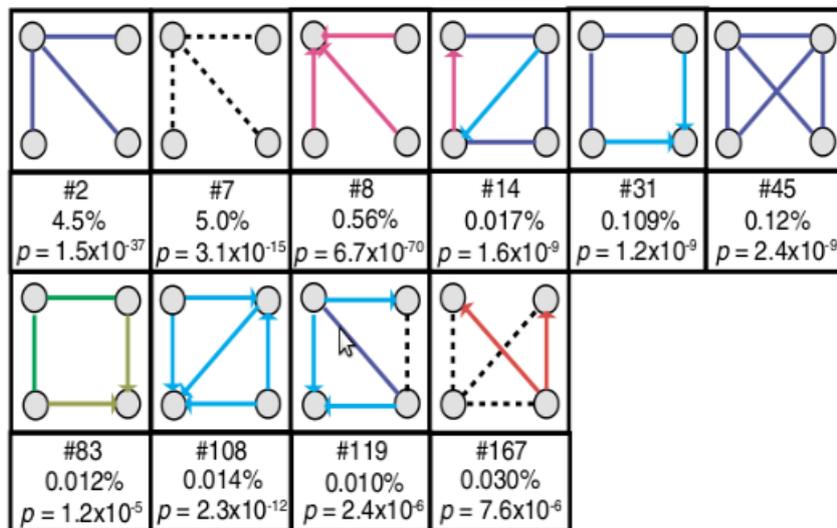


Figura 6: motifs de 4 nós (4n-motifs)

Nota-se que nós específicos (gene perturbado) frequentemente aparecem repetidamente entre os numerosos exemplos de um motif específico. Isto sugere que as instâncias de motifs são unidades estruturais de uma sub-rede de um único motif maior. Tais sub-redes podem destacar as principais perturbações que contribuem para um motif, e mostrar a organização em grande escala de instâncias do motif. A Figura 7 mostra um exemplo de sub-rede de motif único.

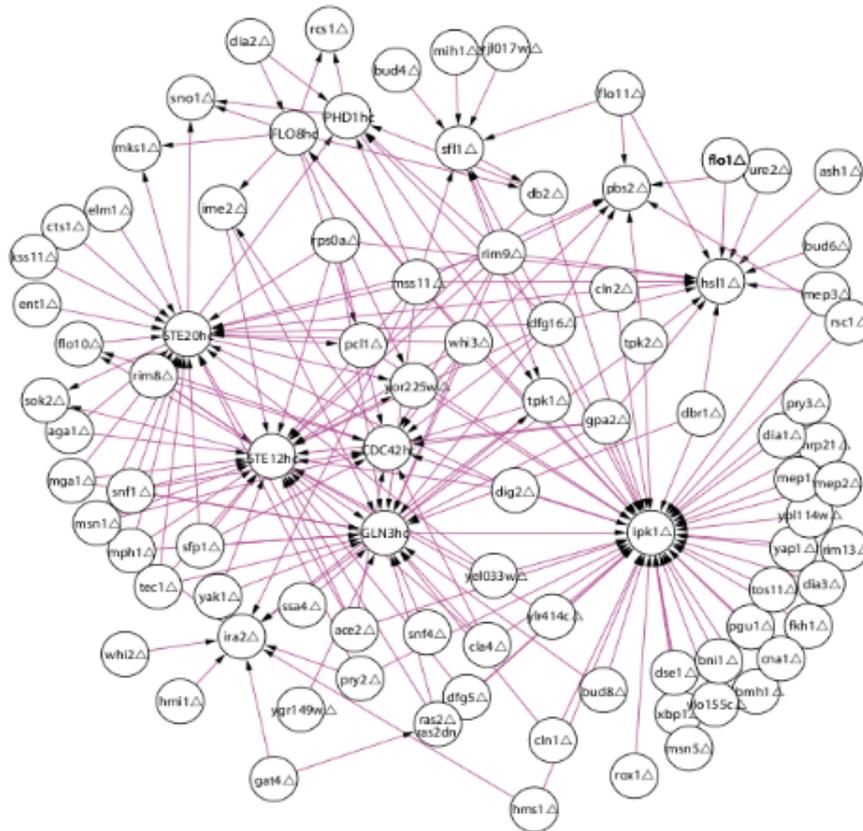


Figura 7: sub-rede de motif único

Na figura 7 identificamos o motif #9 dos 3n-motifs, cujo a modalidade de interação é epistática. Em uma interação epistática, o fenótipo do mutante é o mesmo que um dos dois gene perturbado. Desta forma, as interações epistáticas foram usados para ajudar a identificar e delimitar os fluxos de informação dirigida em sistemas bioquímicos. Conforme mostra a Figura 7, a rede motif epistático é organizada em torno de seis nós principais (hubs). Estas interações repetidas sugerem pontos críticos de fluxo de informação e os genes cujas influências são susceptíveis de fluir através deles.

5 Motifs na música

Motivos (os motifs musicais) aparecem naturalmente em estudos sobre música [7], a palavra motif foi utilizada inicialmente o contexto musical e foi cunhada pelo crítico alemão Hans Wolzogen. Em música, um motif, ou motivo em português, é uma pequena seqüência musical cujo padrão tende a se repetir durante a execução da música.

O famoso motif da quinta sinfonia de Beethoven (sol sol sol mi / fá fá fá ré também identificada pela seqüência de notas musicais da figura 8(b)) tende a identificar a obra, um motif é conectado com um personagem, lugar ou idéia da obra musical, e é executado assim que o sujeito em questão tenha destaque, representando uma importante parte para o enredo. Os motifs aparecem repetidamente

nas composições, sendo em sua forma original, tanto em sua forma original, tanto em variações.

Do ponto de vista das redes complexas, alguns trabalhos recentes estudaram redes formadas a partir de diagramas musicais. Em [8] as redes foram utilizadas para agrupar e classificar diversas músicas com seus gêneros musicais. No trabalho [9] é proposto um algoritmo que constrói redes de acordo com como as notas são conectadas cronologicamente durante a obra musical. A presente seção é baseada no trabalho [7] que utiliza este algoritmo para descrever as redes a partir das músicas. Tal trabalho tem como objetivo a composição artificial de músicas utilizando os motivos detectados a partir das redes musicais.

O algoritmo apresentado em [8] para a formação das redes pode ser entendido resumidamente ao representar as notas musicais como os nós da rede complexa, enquanto a sucessão temporal da nota é descrita como uma ligação. A figura 8(b) exemplifica um motivo formado a partir das notas e seu correspondente sub-grafo. O primeiro nó da rede é a representação da nota semínima pontuada, enquanto sua ligação com a nota colcheia indica a seqüência semínima pontuada para colcheia. O segundo nó tem uma ligação para o nó representativo da semínima, enquanto que este tem outra ligação com a colcheia, evidenciando que a colcheia ocorre duas vezes na seqüência.

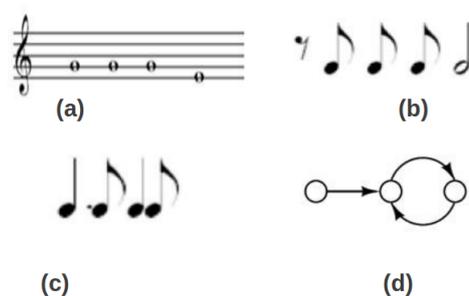


Figura 8: Exemplo de motivo musical tonal(a) e rítmico(b), sendo que esta representa o motivo da quinta sinfonia de Beethoven, três semínimas em seqüência, seguida de uma mínima. (c) representa uma seqüência musical de semínima pontuada, colcheia, semínima e colcheia, enquanto que (d) apresenta a rede formada por esta seqüência

Devido às características musicais são produzidas duas redes distintas, uma rede tonal cuja representação musical pode ser vista na figura 8(a), e uma rede rítmica cuja representação é evidenciada em 8(b). Como fonte musical foi utilizada uma música folclórica japonesa, cuja tradução fonética é Sato-no-aki. O autor não evidencia qual foi o método utilizado para a detecção dos motivos, porém redes formadas pelo método em questão a partir de músicas não geram redes que chegam à ordem de com nós, fazendo com que mesmo métodos de “força bruta” consigam identificar os motivos.

Para a composição das músicas geradas artificialmente utiliza-se um motivo tonal e um rítmico aleatoriamente, formando assim combinações inexistentes na música original, porém ainda agradáveis a ouvintes humanos. No caso da música Sato-no-aki, foram gerados como resultados 13 motivos tonais e 11 motivos rítmicos. Levando em conta restrições musicais quanto a combinações destes elementos, foram gerados ao todo 40 pares de motivos, sendo 16 deles contendo 2 nós, 16 contendo 3 nós, 4 contendo 4 e 2 contendo 5 nós. A Figura 9 exemplifica os motivos tonais resultantes.

Todos os arquivos musicais podem ser encontrados no website <http://cktse.eie.polyu.edu.hk/MUSIC/motifs>.

O autor por fim conclui que é possível identificar uma música utilizando apenas um motivo, assim como identificar mensagens semânticas presentes na obra musical completa. É evidenciado que o tratamento a partir de motivos de redes pode ser utilizado, assim como é possível a produção artificial de músicas agradáveis utilizando apenas combinações aleatórias de redes de motivos tonais

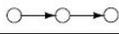
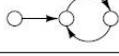
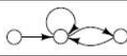
Name	Rhythmic motif	Network motif	Frequency
R ₁			7
R ₂			25
R ₃			7
R ₄			6
R ₅			6
R ₆			12
R ₇			4
R ₈			6
R ₉			7
R ₁₀			4
R ₁₁			6
R ₁₂			5
R ₁₃			4

Figura 9: Tabela contendo motifs em música

e rítmicos.

6 Motifs sobre redes de co-autoria popular: wikipedia

A Wikipédia ([HTTP://wikipedia.com](http://wikipedia.com)) – uma enciclopédia de domínio público on-line, recebeu grande atenção quando um artigo publicado na revista Nature afirmou que o número médio de erros na Wikipédia é menor que a enciclopédia Britânica, sendo esta uma respeitada enciclopédia tradicional. Tal fato deve-se à estrutura da Wikipédia, onde qualquer usuário pode editar um verbete, tornando possível que qualquer usuário contribua com seu conteúdo. Seu conceito pode ser chamado de sabedoria das massas – com uma grande quantidade de usuários acessando seus verbetes é provável que algum deles tenha alguma informação relevante a acrescentar, tornando o verbete mais completo à medida que ele é requisitado.

Entretanto pode-se esperar que a quantidade de edições de um verbete, e assim sua qualidade esteja relacionado com sua quantidade de visitas, e também relacionado com o público que mantém o verbete. Para estudar a relação entre páginas e autoria, [10] propôs um método baseado no uso de Motifs das redes complexas formadas a partir da relação entre autores e páginas da Wikipédia.

Para tal estudo foram escolhidas duas classes de verbetes na Wikipédia, os verbetes relacionados com sociólogos famosos, sendo estes subdivididos entre os franceses, britânicos, alemães, americanos, italianos e os nascidos no século XVIII, com os verbetes de jogadores de futebol ingleses, sendo subdivididos em jogadores do Chelsea, Everton e West Ham United. Todas as informações foram extraídas da versão em língua inglesa ad Wikipédia.

Para construir a rede optou-se por definir dois tipos distintos de nós, os de autoria e o nó página. Um nó página representa um verbete dentro da Wikipédia, enquanto um nó autor representa um autor que participou da edição deste verbete. Uma ligação entre um nó autor e um nó página será

feita caso aquele autor editou aquela página. Uma ligação entre dois nós páginas será dada caso exista hiperligação entre elas. Não é possível ligação entre nós autor. Todos os nós utilizados foram indirecionados.

Foram gerados em seguida uma lista com todos os possíveis motifs possíveis, excluindo aqueles que não possuíam nós autor – o interesse deste estudo é determinar a relação entre autores nos dois grupos selecionados. Foram considerados motifs de dois a cinco nós, resultando assim em 134 motifs possíveis.

Ao construir as redes foi utilizado o software PageAnalyzer (<http://github.com/ChrisSalij/PageAnalyzer>), um software crawler que percorre a Wikipédia fornecendo as informações de autoria e hiperligação das páginas definidas. Para construir as redes foi definido que apenas usuários registrados seriam considerados, assim como seriam desprezados robôs (bots) de edição. Para a procura dos motifs dentro das redes foi utilizado o método GraphGrep [11], que faz a procura dos motifs dentro da rede já considerando possíveis automorfismos (motifs que são completamente equivalentes aos outros mesmo tendo rótulos diferentes).

Devido à problemas ao considerar redes de tamanhos distintos, optou-se por comparar duas redes de tamanho semelhante, a rede formada pelos jogadores do *West Ham United*, e a rede formada pelos sociólogos alemães. Para comparar as duas redes optou-se por utilizar a grandeza $\log(\frac{f_i}{s_i})$, onde f_i e s_i são as freqüências relativas de aparecimento do motif nas redes de jogadores de futebol, sociólogos respectivamente. Tal grandeza é apresentada na figura 10.

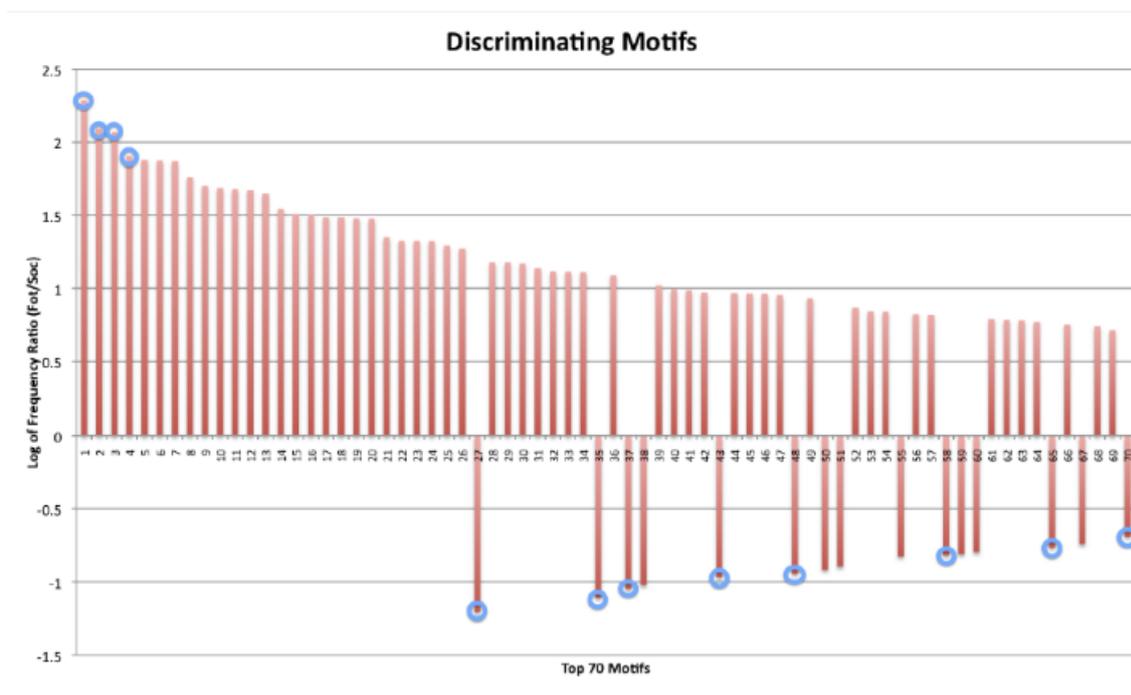


Figura 10: Histograma do logaritmo da razão entre as freqüências relativas dos motifs formados pelas redes do West Ham United e dos sociólogos alemães. Os motifs marcados com círculos azuis são mostrados na figura 11

Devida a forma da grandeza medida, os motifs que tem valor positivo no gráfico são mais presentes na rede de jogadores de futebol, enquanto que valores negativos apresentam motifs mais freqüentes na rede de sociólogos. A figura mostra os motifs [12] mais freqüentes nas duas redes.

Ao analisar os motifs mais freqüentes pode-se notar que no caso dos jogadores de futebol é

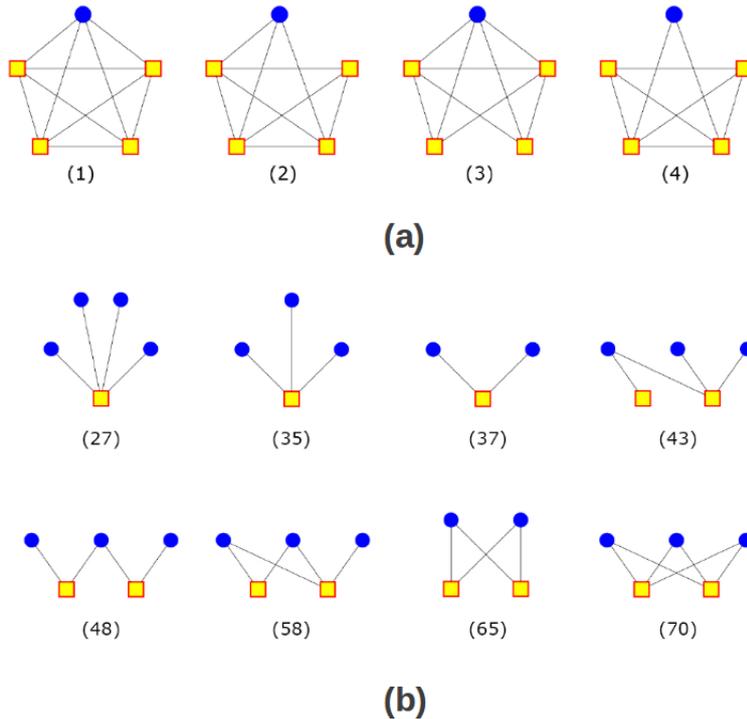


Figura 11: Motifs mais presentes na rede de jogadores de futebol(a) e na rede de sociólogos (b). Nós autos são descritos como círculos azuis, enquanto que nós página são quadrados amarelos

muito comum que um único autor participa da edição de diversas páginas, enquanto que no caso dos sociólogos diversos autores contribuem para um único verbete. Desta forma mostrou-se que avaliar os motifs gerados por ambas as redes pode de fato caracterizar o tipo de autoria presente nos verbetes.

7 Conclusão

Neste estudo, aonde foram resumidos três trabalhos referentes a aplicações dos métodos para enfrentar os desafios da análise de motifs em redes complexas, mostramos que a análise de tais sub-estruturas pode desvendar informações sobre a rede como um todo. Tais informações, as vezes, passam despercebidas pois nem sempre estão sendo apresentadas de uma forma que possamos interpretá-la em sua totalidade. Sendo os motifs uma poderosa ferramenta, que combinadas com alguns conhecimento prévios sobre a rede, podem ser utilizados para caracterizar uma redes complexas. Para isto existem diversas abordagens e métodos de identificação de motifs em redes. Especificamente, a maior parte dos métodos utilizam técnicas estatísticas para identificar padrões significantes em vários tipos de redes. Utilizando uma rede randomizada e a hipótese nula da rede, os padrões que ocorrem com maior frequência do que esperado ao acaso pode ser identificado. Estes motifs que aparecem com alta frequência podem informar padrões da rede que sejam relevantes. No caso da rede de interações genéticas, a união de todas as instâncias de um motif específico pode ilustrar a distribuição deste motif dentro da rede completa, permitindo a identificação de todos os genes envolvidos em tais interações e pode destacar os genes que dominam a ocorrência de um determinado motif. Já no caso da música os motifs são utilizados na geração de músicas artificialmente, onde combina-se de forma aleatória um motif tonal e um motif rítmico, formando assim combinações inexistentes na música original, mas ainda agradáveis aos ouvidos humano. Ainda

falando do campo musical, o trabalho mostrou que é possível identificar uma música utilizando apenas um motif. A análise dos motifs na estrutura de páginas da wikipédia pertiu, através de duas redes distintas, ou seja, redes formadas com a estrutura de dois diferentes assunto da wikipédia, identificar padrões que podem a princípio serem utilizados para identificar qual tipo de rede que está se tratando. Como no caso da rede de jogadores e sociólogos, foi possível mostrar que é muito comum que um único autor participe da edição de diversas páginas, enquanto no caso da rede de sociólogos ocorre que muitos autores contribuem em um única página.

Referências

- [1] NEWMAN, M. E. J. *The structure and function of complex networks*. SIAM Reviews, v. 45, n. 2, p. 167 – 256, 2003.
- [2] Artigo Milo
- [3] Rosenfeld N, Elowitz MB, Alon U (November 2002). *Negative autoregulation speeds the response times of transcription networks*. J. Mol. Biol. 323 (5): 785–93. doi:10.1016/S0022-2836(02)00994-4. PMID 12417193.
- [4] N. Kashtan, S. Itzkovitz, R. Milo and U. Alon *Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs*, Vol. 20 no. 11 2004, pages 1746–1758 doi:10.1093/bioinformatics/bth163
- [5] R James Taylor, Andrew F Siegel and Timothy Galitski *Network motif analysis of a multi-mode genetic-interaction network*, Published: 2 August 2007 Genome Biology 2007, 8:R160 (doi:10.1186/gb-2007-8-8-r160)
- [6] Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila Campillo I, Shannon P, Galitski T: *Derivation of genetic interaction networks from quantitative phenotype data*. Genome Biol 2005, 6:R38.
- [7] Can Yang, Chi K Tse and Xiaofan Liu: *Analyzing and Composing Music from Network Motifs*
- [8] CORREA, D. C. ; SAITO, J. H. ; COSTA, L. da F. . *Music genres: beating to the rhythms of different drums*. New Journal of Physics, v. 12, p. 1-10, 2010.
- [9] C.K. Tse, X. Liu and M. Small, *Analyzing and composing music with complex networks: finding structures in Bach's, Chopin's and Mozart's*, Int. Symp. Nonlinear Theory and Its Applications, udapest, Hungary, pp. 5–8, September 2008.
- [10] Página da Wikipédia: <http://pt.wikipedia.org>
- [11] R. Giugno and D. Shasha. *GraphGrep: A fast and universal method for querying graphs*. In International Conference on Pattern Recognition, volume 16, pages 112–115, 2002.
- [12] Guangyu Wu, Martin Harrigan, Pádraig Cunningham: *A Characterization of Wikipedia Content Based on Motifs in the Edit Graph*. School of Computer Science & Informatics, University College Dublin, Technical REport UCD-CSI-2011-02.