

Universidade de São Paulo  
Campus Butantã  
Instituto de Matemática e Estatística

UM TESTE DO ALGORITMO DE MODULARIDADE LOUVAIN  
COMO UMA FERRAMENTA PARA DETECTAR ESPÉCIES-CHAVE  
EM REDES DE INTERAÇÕES MULTICAMADA

Aluno: Henrique Suzuki Requejo  
Orientador: Prof. Dr. Marco A. R. Mello  
Curso: Matemática Aplicada e Computacional  
Habilitação: Ciências Biológicas

São Paulo, 2020

# **Um teste do algoritmo de modularidade Louvain como uma ferramenta para detectar espécies-chave em redes de interações multicamada**

Trabalho de Conclusão de Curso submetido  
à Universidade de São Paulo, como requisito  
necessário para obtenção do grau de Bacharel  
em Matemática Aplicada e Computacional

São Paulo, 2020

# Conteúdo

<b>1</b>	<b>Resumo</b>	<b>1</b>
<b>2</b>	<b>Modularidade de Newman e Girvan</b>	<b>2</b>
<b>3</b>	<b>Método Louvain para otimização de modularidade</b>	<b>4</b>
<b>4</b>	<b>Modularidade Multicamada</b>	<b>7</b>
<b>5</b>	<b>Parâmetros de acoplamento (<math>\omega</math>) e resolução (<math>\gamma</math>) na prática</b>	<b>8</b>
<b>6</b>	<b>Variável <math>G</math> e <math>G_{norm}</math></b>	<b>13</b>
6.1	Variável $G$ . . . . .	13
6.2	Variável $G_{norm}$ . . . . .	14
<b>7</b>	<b>Visão geral da rede Morcegos-plantas</b>	<b>16</b>
<b>8</b>	<b>Resultados da rede Morcegos-plantas</b>	<b>17</b>
8.1	Distribuição de $\mathbf{G}_{norm}$ . . . . .	17
8.2	Variação de $\overline{\mathbf{G}}$ por $\omega$ . . . . .	18
8.3	Seleção das espécies com maior $\mathbf{G}_{norm}$ . . . . .	22
<b>9</b>	<b>Visão geral da rede Formigas-planta</b>	<b>23</b>
<b>10</b>	<b>Resultados da rede Formigas-planta</b>	<b>24</b>
10.1	Distribuição de $\mathbf{G}_{norm}$ . . . . .	24
10.2	Variação de $\overline{\mathbf{G}}$ por $\omega$ . . . . .	25
10.3	Seleção das espécies com maior $\mathbf{G}_{norm}$ . . . . .	29
<b>11</b>	<b>Comparação entre nós com alto <math>\mathbf{G}_{norm}</math> com nós centrais da rede agregada</b>	<b>30</b>
11.1	Comparação das centralidades com alto $\mathbf{G}_{norm}$ para a rede Morcegos-Plantas	31
<b>12</b>	<b>Discussão preliminar</b>	<b>32</b>
<b>13</b>	<b>Próximas etapas</b>	<b>33</b>
13.1	Outubro . . . . .	33
13.2	Novembro . . . . .	33
<b>14</b>	<b>Referências</b>	<b>34</b>

# 1 Resumo

Análises de modularidade oriundas da ciência de redes têm sido usadas na Ecologia para operacionalizar conceitos como guilda, grupo funcional e papel funcional. Contudo, essas análises têm sido feitas apenas para redes monocamada, ou seja, que contém só um tipo de interação, portanto, apenas uma classe de arestas. No caso do algoritmo de modularidade Louvain, uma das grandes diferenças entre a tradicional versão monocamada e a nova versão multicamada é a constante de acoplamento. Essa constante, implementada no algoritmo multicamada, torna possível que um mesmo nó pertença a dois módulos diferentes em camadas diferentes. Neste projeto, queremos identificar os nós que resistem ao aumento dessa força de acoplamento entre camadas, permanecendo em mais de um módulo em diferentes camadas ao aumentarmos essa força de acoplamento. Em redes de interações ecológicas, esses nós resistentes devem representar espécies que desempenham papéis funcionais importantes. Por exemplo, animais que são importantes tanto para a polinização, quanto para a dispersão de sementes de duas ou mais famílias de plantas. Para classificar as espécies com relação a sua importância, comparamos a curva do número de módulos aos quais cada nó pertence em função da variação na constante de acoplamento e do parâmetro de resolução. Assim, selecionaremos os nós que apresentam menor decaimento de módulos, ou seja, que permanecem um mais de um módulo mesmo com força de acoplamento entre camadas, considerando diferentes cenários de resolução. Até o momento foram analisadas as redes de interações entre morcegos e plantas e de formigas e plantas. A próxima etapa é avaliar se ganhamos informação relevante, tanto no nível do nó como no nível dos módulos, na abordagem multicamada em comparação à monocamada. Em seguida, repetiremos essas mesmas análises para redes sintéticas, dessa maneira, poderemos verificar se o método é eficaz em várias redes diferentes. Isso nos permitirá entender melhor o que faz as espécies que eles representam serem mais importantes do que outras para manter funções e serviços ecossistêmicos vitais.

## 2 Modularidade de Newman e Girvan

A ciência de redes tem sido usada para investigar sistemas ecológicos desde o século XIX (Ings & Hawes, 2018). Dentre vários avanços, essa abordagem nos permitiu entender melhor como várias espécies afetam umas às outras por meio de interações diretas e indiretas. Esse entendimento levou a importantes insights sobre a estrutura, função e dinâmica de redes de interações (Pilosof et al., 2017).

Uma forma de obtermos estes insights é olhar para a estrutura de módulos da rede, o que pode revelar alguns padrões escondidos em meio à complexidade de nós e conexões. Estes módulos são compostos por nós mais conectados entre si do que com outros nós de fora do módulo. A capacidade de encontrar e analisar esses módulos pode fornecer ajuda inestimável para entender mais a fundo a estrutura das redes (Newman & Girvan, 2004).

É relativamente fácil identificar módulos em redes pequenas e pouco conectadas. Porém, essa tarefa se torna muito difícil quando analisamos redes mais complexas, com milhares, ou até mesmo milhões, de nós distribuídos em várias camadas. Então, como podemos identificar módulos em redes complexas? Existem muitas formas, como por exemplo métodos aglomerativos, divisivos, baseados em cliques e de otimização de modularidade. Neste projeto vamos focar na otimização de modularidade.

Para otimizarmos a modularidade, precisamos de uma métrica que indique se uma específica divisão da rede em módulos é melhor do que outra, para então buscar qual seria a melhor divisão de módulos possível. Uma métrica para medir a qualidade da modularidade de redes foi definida pela primeira vez por Newman e Girvan (Newman & Girvan, 2004) como:

$$Q = \sum_{i=1}^k (e_{ii} - a_i^2) \quad (1)$$

Onde  $e_{ii}$  representa a fração das conexões observadas dentro do módulo  $i$  em relação ao total de conexões dentro da rede e  $a_i^2$  representa a quantidade esperada de conexões no módulo  $i$ .

Para determinar a proporção de conexões dentro dos módulos, suponha que temos uma possível divisão de módulos dentro da rede representada pela matriz de adjacência  $A_{ij}$ , seja  $C_i$  o módulo que o nó  $i$  está contido, então, a fração de conexões que conectam os vértices que estão dentro de um mesmo módulo é:

$$\frac{\sum_{ij} A_{ij} \delta(c_i, c_j)}{\sum_{ij} A_{ij}} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, c_j) \quad (2)$$

Onde  $\delta$  é o delta de kronecker, que retorna o valor 1 caso  $c_i = c_j$  e zero caso contrário; a variável  $m$  representa o número total de conexões da rede. Falta agora definirmos uma forma de medir a quantidade esperada de nós dentro do módulo. Se conectarmos os vértices de forma aleatória, preservando os graus de cada nó, representados pela variável  $k$ , temos que a probabilidade de existir uma conexão entre os nós  $i$  e  $j$  é  $k_i k_j / 2m$ . Dessa forma, temos que a modularidade descrita acima pode ser reescrita de uma forma mais operacional (Blondel et al., 2008) como:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j) \quad (3)$$

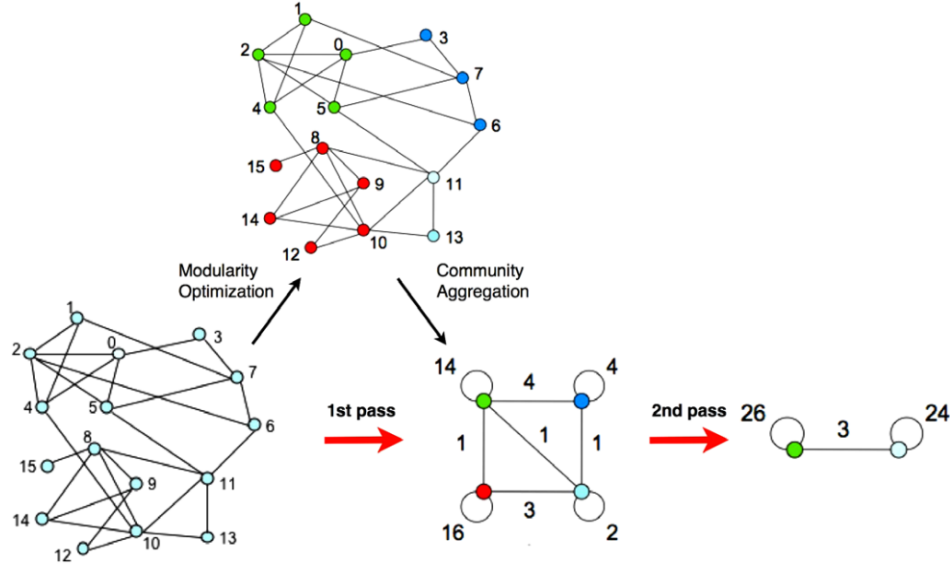
Os valores da modularidade descrita acima podem variar entre zero e 1.

### 3 Método Louvain para otimização de modularidade

Se dividirmos a rede em módulos e o valor de modularidade definido por Newman e Girvan for alto, temos que a divisão de módulos escolhida é boa. Dessa forma, basta dividirmos a rede em diferentes módulos visando maximizar o valor da modularidade Newman-Girvan. Porém, esta não é uma tarefa fácil, já que a otimização de modularidade é um problema NP-difícil (Brandes et al., 2008). Um problema NP (Non Deterministic Polynomial Time) é um problema que leva um tempo exponencial para ser resolvido, por exemplo  $O(2^N)$ . Estes problemas NP demoram muito para serem computados, por exemplo, caso o problema possua uma complexidade exponencial  $O(2^N)$  e  $N$  (tamanho da entrada) seja igual a 100, o programa demoraria um tempo maior que a idade do universo para ser resolvido. Um problema NP-difícil é um problema que é ao menos tão difícil quanto um problema NP-completo.

Esta dificuldade computacional em dividir redes em módulos significativos motivou a procura por métodos heurísticos para encontrar agrupamentos que apresentam um alto valor de modularidade com uma complexidade razoável. Este é o caso do método Louvain (Blondel et al., 2008), que parece funcionar em um tempo  $O(M)$ , onde  $M$  representa o número total de conexões da rede, ou seja, é resolvido em tempo polinomial, neste caso, linear com relação ao número de nós da rede. Essa complexidade computacional permite que redes de grande escala possam ser analisadas em tempo razoável, além de ser simples e de fácil implementação.

A implementação do algoritmo Louvain consiste de dois passos básicos, descritos em detalhe por Blondel et al. (2008). O primeiro passo agrupa os nós em módulos de forma individual, favorecendo otimizações locais de modularidade. O segundo passo transforma a rede original em uma nova rede simplificada, cujo os novos nós são os módulos do passo anterior. As conexões entre estes novos nós correspondem à soma das conexões entre os módulos e cada nó agora tem um laço (conexão que liga um nó a ele mesmo) cujo o peso é a soma das conexões dentro do módulo correspondente. Esses dois passos são repetidos até que não exista mais aumento no valor da modularidade. A figura 1 ilustra o algoritmo.



**Figura 1:** Extraído de Blondel et al. (2008). Visualização dos passos do algoritmo Louvain. Cada passo é feito em duas fases: uma onde a modularidade é otimizada permitindo somente mudanças locais de módulo e uma onde os módulos encontrados são agregados e uma nova rede de módulos é construída.

Apesar deste ser um algoritmo ganancioso, ele apresenta boa detecção de módulos que possuem sentido quando aplicados em redes reais (Blondel et al., 2008; Meunier, 2009; Zhang et al., 2010) e é mais rápido que outros algoritmos similares, como mostrado na tabela 1, extraída de Blondel et al. (2008). Outra vantagem do método é que não precisamos saber previamente o número de módulos que a rede deve ser dividida.

**Tabela 1:** Extraído de Blondel et al. (2008). Comparação entre os tempos computacionais dos algoritmos Clauset et al (CNM), Pons e Latapy (PL), Wakita e Tsurumi (WT) e Louvain (LV). Os resultados mostram a modularidade obtida e o tempo computacional para cada rede/método. Células em branco representam tempos computacionais maiores que 24 horas.

	Karate	Arxiv	Internet	Web nd.edu	Telefone	Web uk2005	WebBase 2001
Nós / Conexões	34/77	9k/24k	70k/351k	325k/1M	2.04M/5.4M	39M/783M	118M/1B
CNM	0.38/0s	0.772/3.6s	0.692/799s	0.927/5034s	-	-	-
PL	0.42/0s	0.752/3.3s	0.729/575s	0.895/6666s	-	-	-
WT	0.42/0s	0.761/0.7s	0.667/62s	0.898/248s	0.553/367s	-	-
LV	0.42/0s	0.813/0s	0.781/1s	0.935/3s	0.76/44s	0.935/3s	0.984/152min

Por outro lado, o método Louvain também apresenta limitações (Good et al., 2010). Dentre as mais expressivas estão o limite de resolução, o problema de degeneração e o fato de o algoritmo não ser determinístico. O problema de limite de resolução ocorre pois o algoritmo pode não parar nos módulos “intuitivos” devido a uma segunda passagem pela modificação dos módulos. Já o problema de degeneração ocorre pois podem existir muitas



soluções com valor de modularidade próximos do máximo global, nesses casos é muito difícil encontrar o máximo global e de realmente afirmar se o máximo global possui significância científica maior que os máximos locais com valores de modularidade similares.

Uma característica interessante do método é que não é possível, pela própria natureza do algoritmo, que um mesmo nó esteja presente em diferentes módulos, essa sobreposição de módulos é chamada dentro da área de modularidade de redes de *overlapping communities*. Porém, ao adaptarmos a forma de calcular a modularidade para multicamada (Mucha et al., 2010), existe a possibilidade de que um mesmo nó pertença a diferentes módulos em diferentes camadas ao ajustarmos um parâmetro de entrada. Esta diferença é interessante e será usada para tentar encontrar espécies-chave que são boas conectoras entre grupos diferentes em camadas diferentes.

## 4 Modularidade Multicamada

Apesar do amplo estudo de redes para compreensão de sistemas e tomada de decisão, estas redes são normalmente estudadas de forma desconexa de outras redes ou então agregando-se várias redes em uma única rede (Piloso et al., 2017). Uma maneira de aumentarmos as informações usadas no estudo de redes é usar redes multicamada (Piloso et al., 2017). Quando usadas para representar sistemas ecológicos, as redes multicamada podem ter suas camadas definidas através de interações de diferentes tipos, ou que acontecem em diferentes localidades ou estações do ano, por exemplo (Piloso et al., 2017).

No caso deste projeto, a rede em foco é a rede multicamada onde as camadas são definidas pelo tipo de interação, ou seja, todos os nós estão presentes em todas as camadas e apenas o tipo de ligação (interações) entre os nós é o que muda em cada camada. Assim, nesse tipo de rede há duas classes maiores de arestas: intracamada, que conecta nós dentro de uma camada, e intercamada, que conecta um mesmo nó entre duas camadas. Este tipo de rede é chamada de *multiplex* (Mucha et al., 2010), que é uma subdivisão da grande categoria definida como redes multicamada.

Redes multiplex também podem ter sua modularidade analisada pelo método Louvain (Mucha et al., 2010). Essa adaptação consiste basicamente na mudança da forma como a função da modularidade é calculada, incluindo um parâmetro  $\omega$  que controla o acoplamento entre as diferentes camadas da rede. Há também o parâmetro de resolução  $\omega$ , que pode enviesar a tendência do algoritmo em encontrar módulos de escalas menores ou maiores. A generalização da equação de modularidade (Mucha et al., 2010) é dada por:

$$Q^M = \frac{1}{\mu} \sum_{i,j,\alpha,\beta} \left[ \left( A_{i\alpha,j\alpha} - \gamma_\alpha \frac{k_{i\alpha} k_{j\alpha}}{2m_\alpha} \right) \delta_{\alpha,\beta} + \omega A_{i\alpha,j\beta} \delta_{ij} \right] \delta(c_{i\alpha}, c_{j\beta}) \quad (4)$$

Onde  $\delta$  é o delta de kronecker, que retorna o valor 1 caso  $c_{i\alpha} = c_{j\beta}$  e zero caso contrário; a variável  $m$  representa a soma de do grau de cada nó na camada  $\alpha$ ;  $A_{i\alpha,j\alpha}$  é a matriz de adjacência na camada  $\alpha$ ;  $A_{i\alpha,j\beta}$  é a matriz de adjacência entre camadas;  $k_{i\alpha}$  representa o grau do nó  $i$  na camada  $\alpha$  e  $\mu = \sum_{i,j,\alpha} A_{i\alpha,j\alpha} + \omega \sum_{i,\alpha,\beta} A_{i\alpha,j\beta}$ . Note que se utilizarmos  $\omega = 0$  e  $\gamma = 1$ , a equação de modularidade generalizada (eq. 4) é proporcional a média da modularidade monocamada (Newman, 2004; Newman & Girvan, 2004) de cada camada. Isto também está descrito em Bianconi (2018), pg. 147-148:

$$Q^M = \frac{1}{\mu} \sum_{\alpha} \frac{1}{2m_\alpha} \sum_{ij} \left( A_{i\alpha,j\alpha} - \gamma_\alpha \frac{k_{i\alpha} k_{j\alpha}}{2m_\alpha} \right) \delta(c_{i\alpha}, c_{j\alpha}) \quad (5)$$

## 5 Parâmetros de acoplamento ( $\omega$ ) e resolução ( $\gamma$ ) na prática

Como a escolha dos parâmetros de resolução e acoplamento é arbitrária, ficam as perguntas: qual seria o melhor valor para aplicar em uma dada rede? Existe apenas um único valor adequado ou o valor a ser escolhido depende do da pergunta que pretendo responder? Quais são os efeitos na prática quando variamos  $\gamma$  e  $\omega$ ? Com o objetivo de responder essas perguntas e auxiliar outros cientistas na escolha dos valores de  $\gamma$  e  $\omega$ , vamos analisar essas questões em detalhes.

O parâmetro  $\omega$  está contido no intervalo (0,1). Zero (0) representa o desacoplamento total das camadas (cada camada é tratada como uma rede individual) e a modularidade final é a média da modularidade de cada camada. Um (1) representa o acoplamento máximo, que faz com que a rede multiplex se comporte como uma rede monocamada. Na prática, isso significa que, quando aumentamos os valores da constante de acoplamento  $\omega$ , aumentamos o peso das arestas intercamada (Mucha et al., 2010), isso faz com que as camadas tenham maior influência umas sobre as outras, favorecendo a formação de módulos multicamada. Quando diminuimos  $\omega$ , o contrário ocorre, o que favorece a formação de módulos monocamada.

O parâmetro de resolução foi introduzido por Reichardt & Bornholdt (2006) para avaliar redes monocamada e depois estendido para a modularidade generalizada multicamada (Mucha et al., 2010). De uma forma geral, se  $\gamma_2 > \gamma_1$ , os módulos encontrados com  $\gamma_2$  possuem menos nós (módulos menores) e são mais numerosos (maior quantidade de módulos). Os módulos encontrados com  $\gamma_2$  podem ser submódulos dos obtidos usando  $\gamma_1$ , mas nem sempre é o caso (Reichardt & Bornholdt, 2006). Lembrando que  $0 \leq \gamma \leq \infty$ . Porém, não faz sentido aumentarmos  $\gamma$  para valores muito altos, já que existe um limite a partir do qual os módulos se tornam tão pequenos que cada módulo passa a ter apenas um nó.

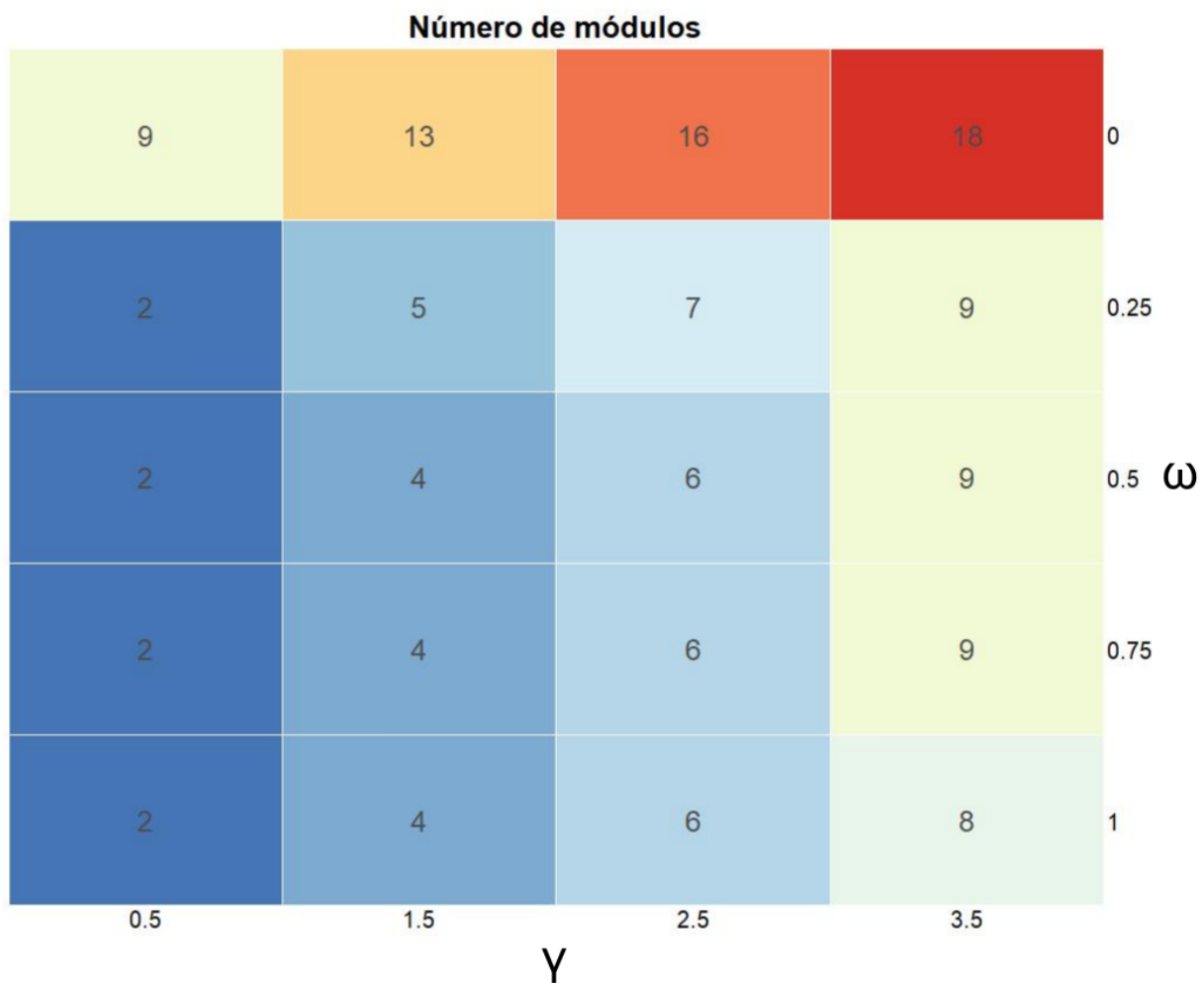
Vamos usar a rede multiplex Famílias de Florença (Kent, 1978) para ilustrar graficamente o que ocorre quando variamos os parâmetros  $\omega$  e  $\gamma$ . Essa rede foi escolhida por ser pequena, o que facilita sua visualização. A figura 2 mostra a distribuição dos módulos sobre as duas camadas da rede para diferentes valores de  $\omega$  e  $\gamma$ .



**Figura 2:** Módulos formados usando diferentes valores de acoplamento  $\omega$  e resolução  $\gamma$ . Valores de  $\omega$  variam no eixo y e valores de  $\gamma$  variam no eixo x.

Quanto maiores o os valores de  $\gamma$  e menores os valores de  $\omega$ , menores e mais numerosos são os módulos e vice-versa. A figura 3 mostra o número de módulos para a rede exemplo

Famílias de Florença, onde podemos ver que ocorre o previsto na teoria.



**Figura 3:** Número de módulos totais da rede para diferentes valores de acoplamento  $\omega$  e resolução  $\gamma$ . Valores de  $\omega$  variam no eixo y e valores de  $\gamma$  variam no eixo x.

Mas, afinal, quais valores devo escolher para esses dois parâmetros? A resposta mais correta é: depende. Depende de qual propriedade da rede queremos focar e qual pergunta queremos responder com esse enfoque. Por exemplo, se quisermos verificar quais são os “grandes módulos” da rede, devemos usar um valor de  $\gamma$  mais baixo, mas se quisermos olhar para os módulos menores (mais “íntimos”) ou submódulos dos “grandes módulos” obtidos com um  $\gamma$  mais baixo, um  $\gamma$  maior seria mais indicado.

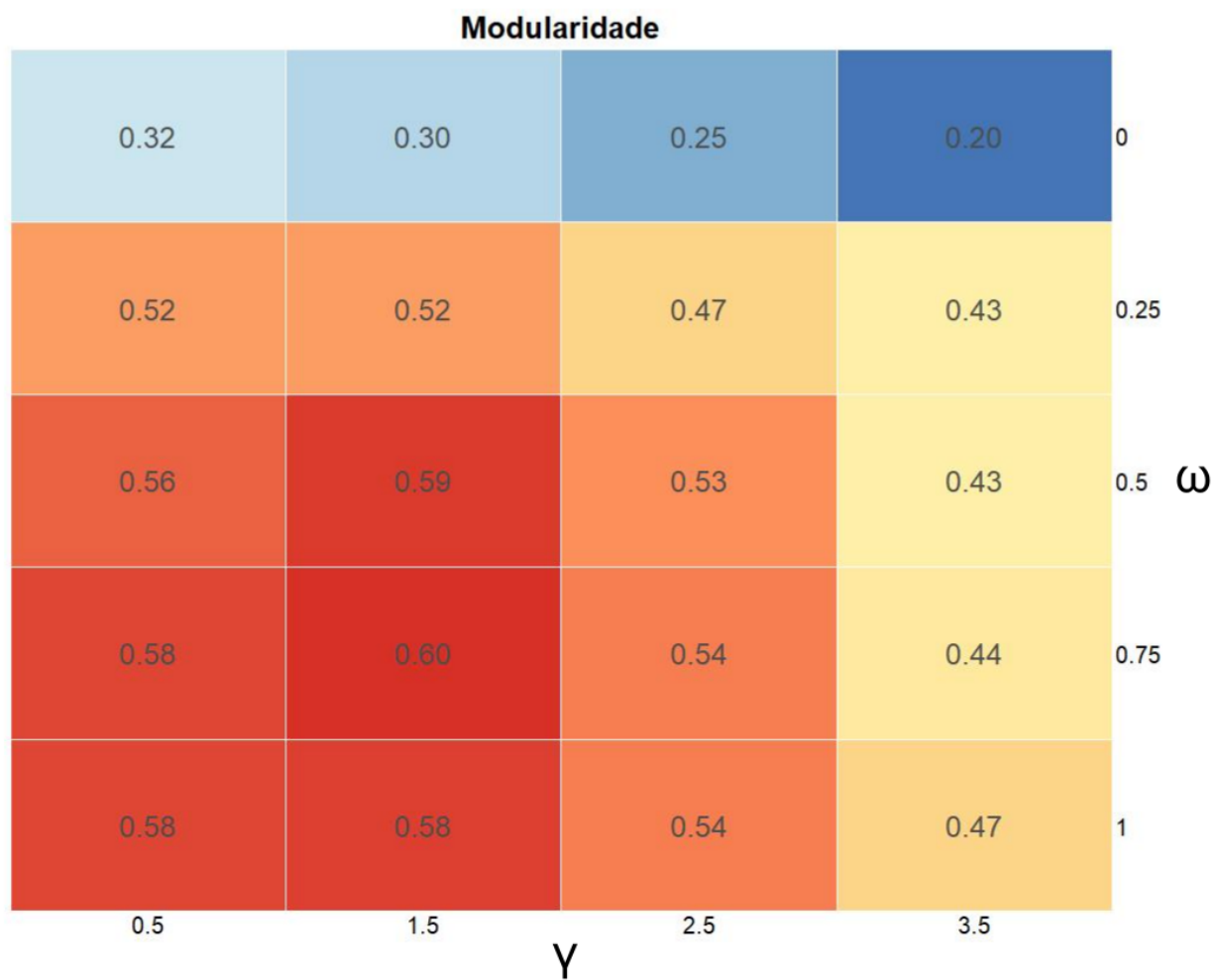
Se quisermos que as conexões de uma camada influenciem mais a outra camada, devemos usar um  $\omega$  mais alto, caso contrário, melhor usar um  $\omega$  mais baixo. Vamos usar agora como exemplo a rede morcego-planta (Mello et al., 2019), onde existem duas camadas: uma de nectarívoria e outra de frugivoria. Se quisermos encontrar módulos onde existam morcegos com dietas mais similares entre si, um  $\omega$  mais alto é recomendado, pois a frugivoria e a nectarívoria ficaram com alta influência uma sobre a outra no cálculo dos módulos. Já se quisermos separar os morcegos com uma preferência maior por flores, por frutos ou que tenham uma dieta equilibrada, podemos usar um  $\omega$  menor, porque assim aumenta a tendência dos módulos se formarem pesando mais as interações dentro de cada camada da rede. Isso faz com que os morcegos estejam em grupos que priorizam mais a nectarívoria ou a frugivoria ou até mesmo que possuem interações equilibradas entre camadas (estão em dois grupos em camadas diferentes).

Interpretação, conhecimento específico da área que a rede representa e saber o que queremos enxergar são os fatores principais para a escolha dos valores de  $\omega$  e  $\gamma$ .

Caso quisermos apenas obter uma distribuição de módulos confiável sem a necessidade de interpretação, podemos escolher um valor de  $\omega$  e  $\gamma$  que maximiza a modularidade. Os valores de  $\omega$  e  $\gamma$  que maximizam a modularidade diferem para cada rede. A figura 4 mostra os valores da modularidade para diferentes valores de  $\omega$  e  $\gamma$  da rede exemplo Famílias de Florença.

Com a resolução do ensaio feito, os valores mais indicados para a rede famílias de florença seriam  $\omega = 0.75$  e  $\gamma = 1.5$ . Para valores mais precisos, basta aumentar as partições de  $\gamma$  e  $\omega$  ou refinar os valores de  $\gamma$  e  $\omega$  em torno dos máximos obtidos no ensaio anterior.

Também existe a possibilidade de que os valores de modularidade fiquem muito próximos uns dos outros. Nesse caso, é comum que os módulos fiquem muito similares entre si, então, é possível escolher qualquer um dos valores de  $\omega$  e  $\gamma$ . Caso os módulos fiquem muito diferentes entre si (um caso mais raro), uma interpretação de um especialista na área que a rede está retratando é fundamental.

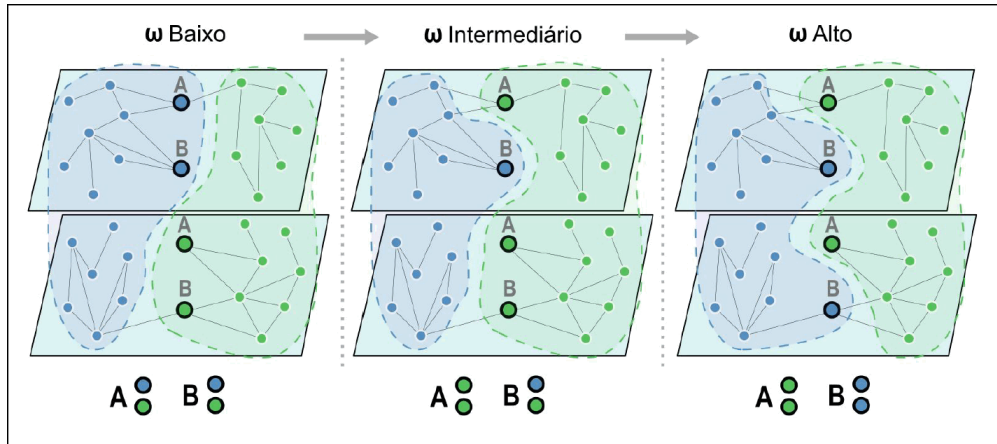


**Figura 4:** Valor da modularidade para diferentes valores de acoplamento  $\omega$  e resolução  $\gamma$ . Valores de  $\omega$  variam no eixo y e valores de  $\gamma$  variam no eixo x.

## 6 Variável $G$ e $G_{norm}$

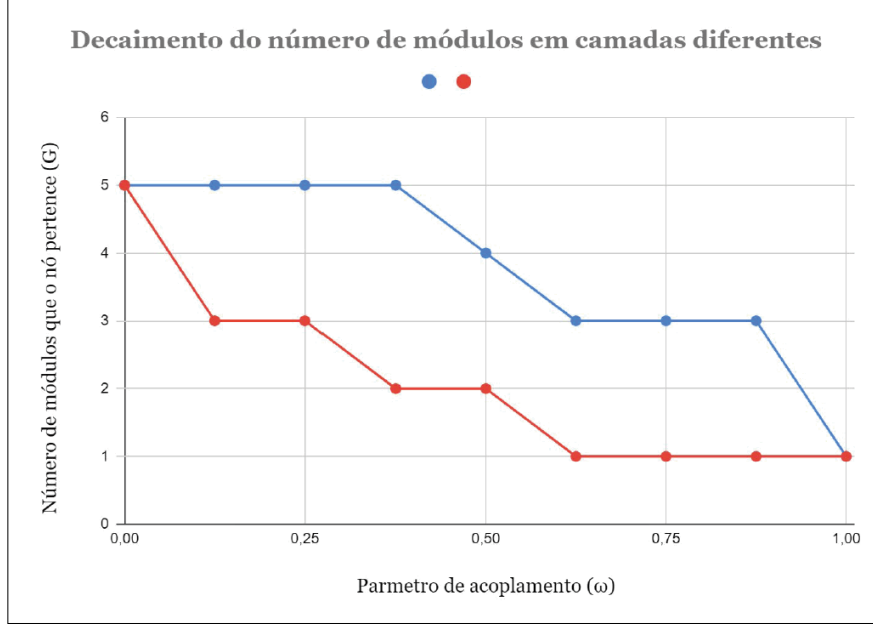
### 6.1 Variável $G$

A variável  $G$  representa o número de módulos que um determinado nó pertence. Como o algoritmo Louvain não permite que um mesmo nó participe de mais de um módulo dentro de uma mesma camada,  $G$  é uma variável discreta inteira que possui o mínimo de 1 e o máximo igual ao número de camadas da rede estudada.  $G$  possui valores bem definidos nos extremos, se  $\omega = 0 \Rightarrow G = 1$  e se  $\omega = 0 \Rightarrow G = \text{número total de camadas da rede}$ . Entre estes valores de  $\omega$  o comportamento de  $G$  é variado, mas tende a cair conforme a força de acoplamento  $\omega$  aumenta. O comportamento do decaimento de  $G$  em relação a  $\omega$  é diferente para cada nó da rede, alguns nós resistem e permanecem em mais de um módulo ao aumentarmos  $\omega$ , já outros rapidamente passam a pertencer a apenas um módulo ao aumentarmos  $\omega$ . A figura 5 ilustra o comportamento de diferentes nós ao aumentarmos  $\omega$ . A figura 6 mostra um exemplo de curva de decaimento de  $G$  em relação a  $\omega$ .



**Figura 5:** Esquema ilustrativo do que ocorre com alguns nós ao aumentarmos o parâmetro  $\omega$ . Para valores de  $\omega$  baixos, vemos que os nós A e B estão em módulos diferentes em camadas diferentes. Ao aumentarmos a força de acoplamento  $\omega$ , o nó A passa a pertencer ao módulo verde tanto na camada superior como na camada inferior, já o nó B resiste e permanece no módulo azul na camada superior e no módulo verde na camada inferior. Se aumentarmos ainda mais a força de acoplamento  $\omega$ , em algum momento o nó B acaba cedendo e passa a pertencer a apenas um módulo nas duas camadas.





**Figura 6:** Exemplo hipotético de decaimentos de  $G$  (módulos que o nó pertence em diferentes camadas) em relação ao parâmetro  $\omega$  (constante de acoplamento) esperados. O nó representado pela cor vermelha possui um decaimento rápido de  $G$ , portanto, representa uma espécie pouco conectiva do ponto de vista da rede multicamada como um todo. O contrário ocorre com a espécie representada pelo nó azul.

Como o algoritmo Louvain não é determinístico (pode retornar diferentes resultados a cada compilação) e a variável  $G$  é inteira e discreta, para termos uma maior confiabilidade precisamos repetir o processo várias vezes para cada valor de  $\omega$  e extrair uma média de  $G$  para cada nó. Assim o valor médio de  $G$  é dado por:

$$\overline{G}_\omega = \frac{1}{I} \sum_{i=1}^I G_i \quad (6)$$

Onde  $I$  se refere ao número de iterações.

Repetindo isso para diferentes valores de  $\omega$ , teremos uma curva de decaimento  $\overline{G}$  por  $\omega$ .

Podemos agora repetir esse processo para diferentes valores de  $\gamma$ , assim vamos obter diferentes curvas  $\overline{G}$  por  $\omega$  para cada valor de  $\gamma$ .

## 6.2 Variável $G_{norm}$

Para facilitar a classificação dos nós de acordo com sua curva  $\overline{G}$  por  $\omega$  por  $\gamma$ , toda a família de curvas será resumida em uma única variável para cada nó, normalizada pela média.

Chamaremos essa variável de  $G_{norm}$ , definida na equação 7.

$$G_{norm} = \frac{\sum_{j=1}^{P_2} \sum_{i=1}^{P_1} \overline{G}_{i,j}}{\frac{1}{N} \sum_{k=1}^N \sum_{j=1}^{P_2} \sum_{i=1}^{P_1} \overline{G}_{i,j,k}} \quad (7)$$

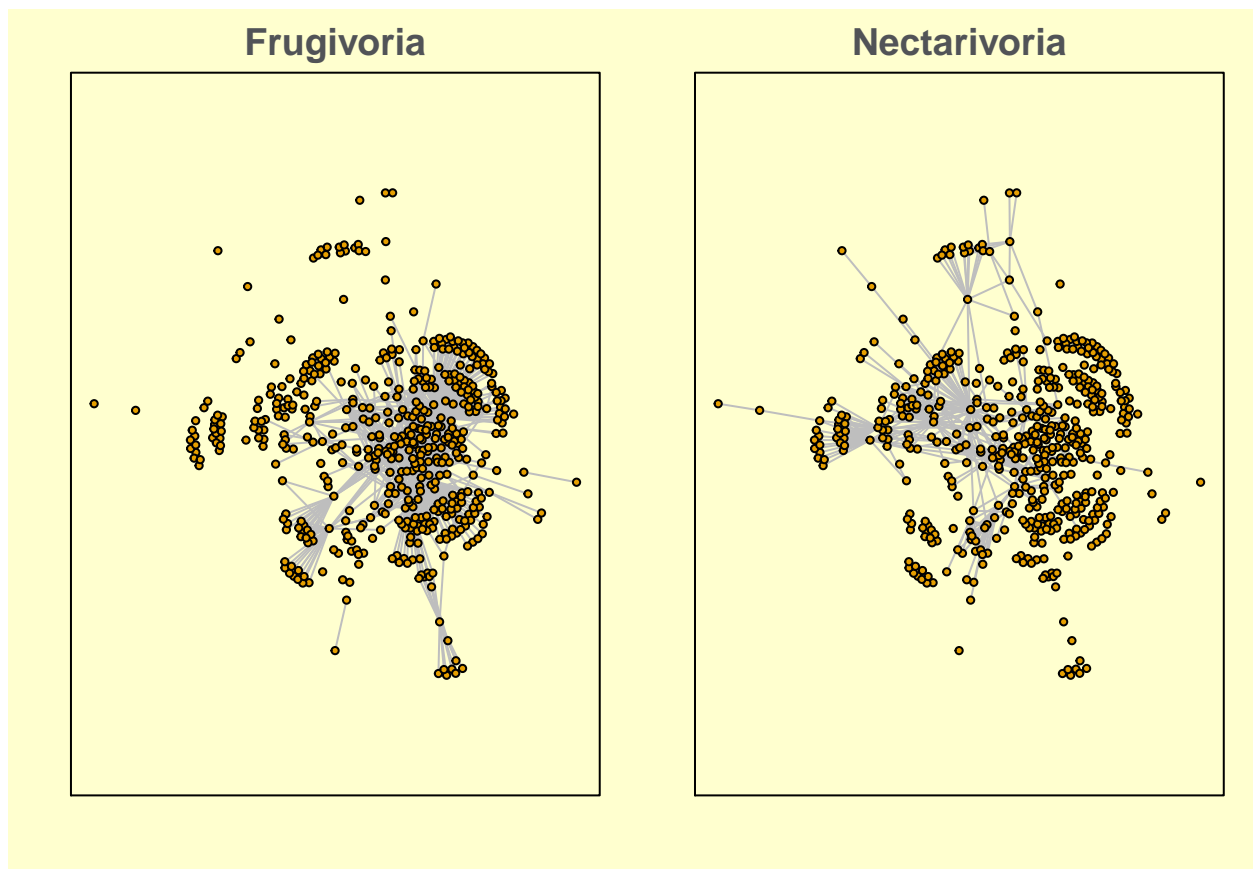
Onde  $P_1$ ,  $P_2$  são o número de partições de  $\omega$  e  $\gamma$ ;  $N$  é u número de nós na rede;  $i$ ,  $j$  são os índices de  $\overline{G}$  dentro de  $\omega$ ,  $\gamma$ ; e  $k$  é o índice do nó na rede (número de identificação do nó).

## 7 Visão geral da rede Morcegos-plantas

A rede Morcegos-plantas (Mello et al., 2019) possui 2 camadas (Nectarivoria e Frugivoria), 512 nós e 1248 conexões. A tabela 2 mostra o resumo das propriedades da rede. A figura 7 apresenta uma visão geral da rede.

**Tabela 2:** Propriedades da rede Morcegos-plantas

Propriedade	Valor
Número de Camadas	2
Tipo de conexões	Nectarivoria e Frugivoria
Número de nós	512
Número de conexões	1248



**Figura 7:** Visão geral da rede Morcegos-plantas .

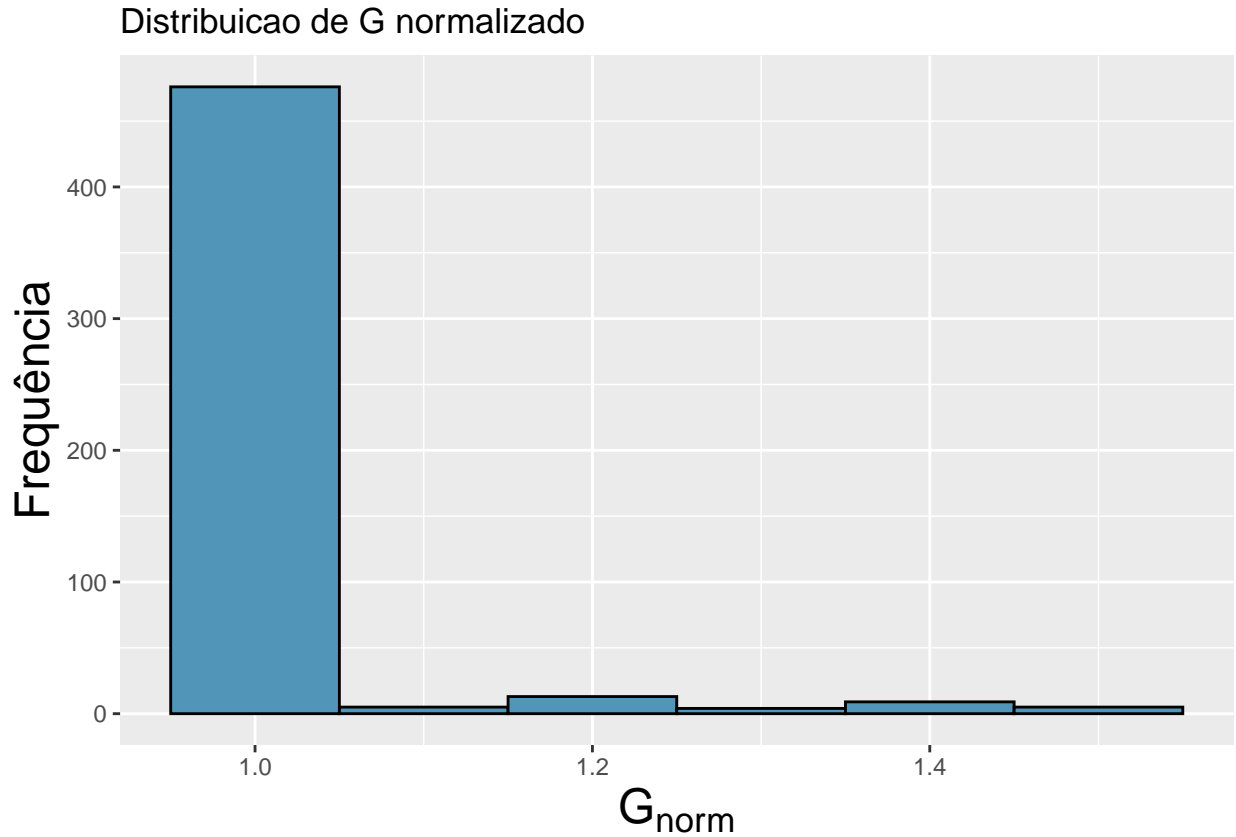
## 8 Resultados da rede Morcegos-plantas

### 8.1 Distribuição de $G_{norm}$

A variável  $G$  foi calculada para 10 partições de  $\omega$ , ou seja, o tamanho do passo dado dentro de  $\omega$  foi de 0.1. O processo foi repetido para 16 partições de  $\gamma$ , com  $\gamma$  começando em 0.25, com passos de 0.25 até um  $\gamma$  máximo de 4. O cálculo de  $\overline{G}$  foi feito usando 100 iterações. A tabela 3 resume os parâmetros de execução do código e a figura 8 mostra a distribuição dos valores de  $G_{norm}$  médio obtidos.

**Tabela 3:** Parâmetros de execucao

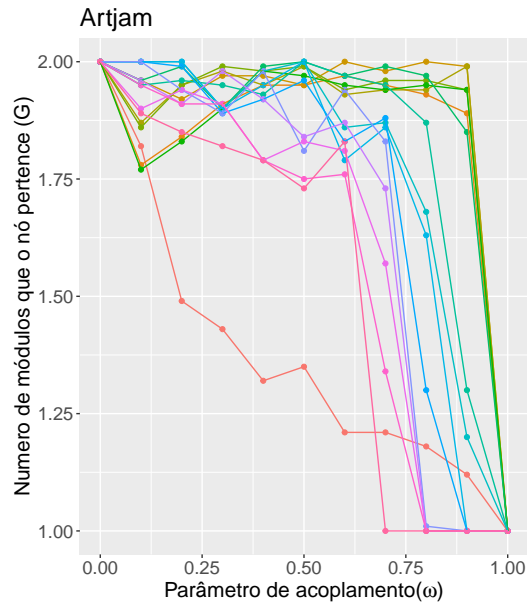
Parâmetro	Valor
Iterações	100
Partições de omega	10



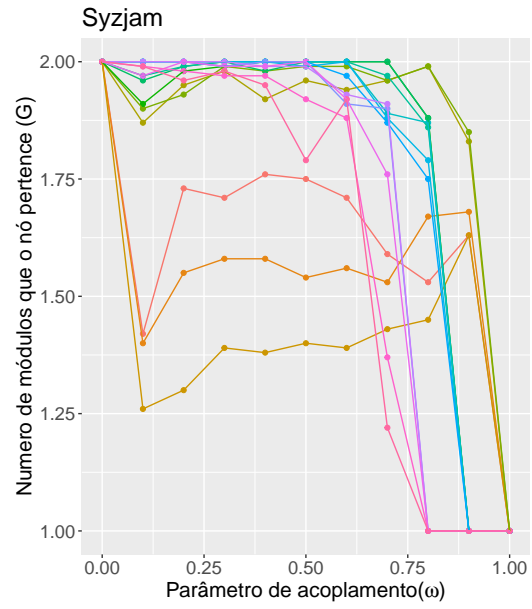
**Figura 8:** Distribuição de  $G_{norm}$  médio da rede Morcegos-plantas .

## 8.2 Variação de $\overline{G}$ por $\omega$

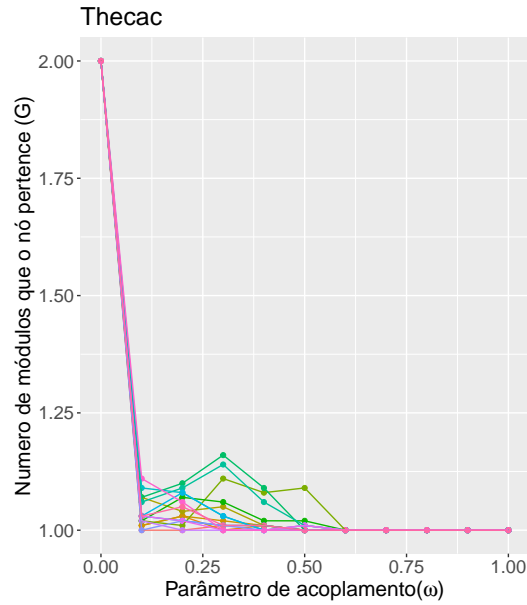
Como temos dados em 3 dimensões  $(\overline{G}, \omega, \gamma)$  temos algumas formas diferentes para apresentar os valores de  $\overline{G}$  em relação a  $\omega$  e  $\gamma$ , não sei dizer se devemos usar uma delas, as três ou alguma outra. A figura 9 mostra curvas de decaimento de  $\overline{G}$  por  $\omega$  para diferentes nós com diferentes valores de  $G_{norm}$  e para diferentes valores de  $\gamma$ . A figura 10 mostra a superfície 3D formada por  $\overline{G}$  em relação a  $\omega$  e  $\gamma$ . A figura 11 mostra a mesma superfície da figura 10 mas no formato de mapa de calor.



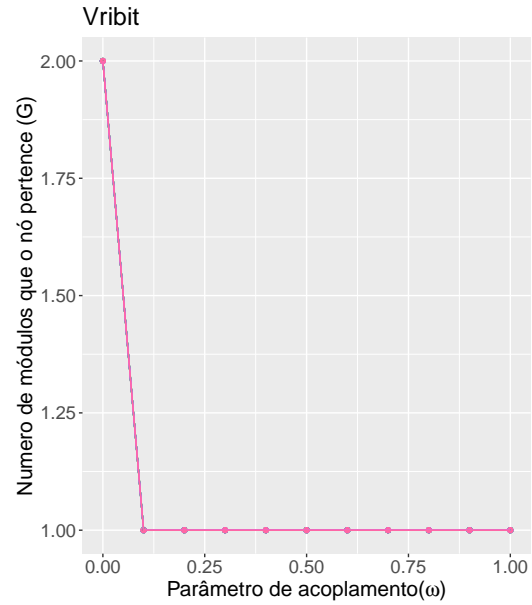
(a) Artjam .  $G_{norm} = 1.548$



(b) Syzjam .  $G_{norm} = 1.538$

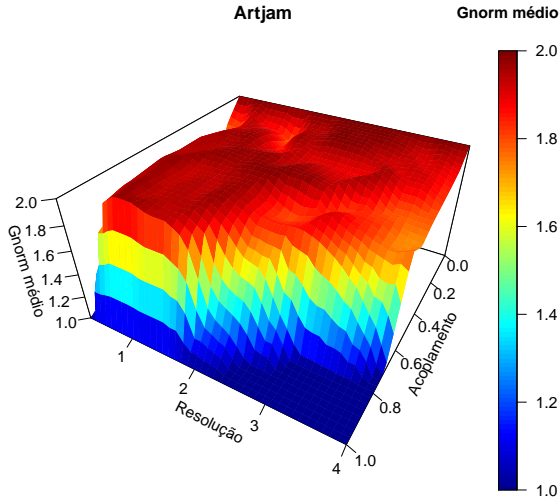


(c) Thecac .  $G_{norm} = 0.990$

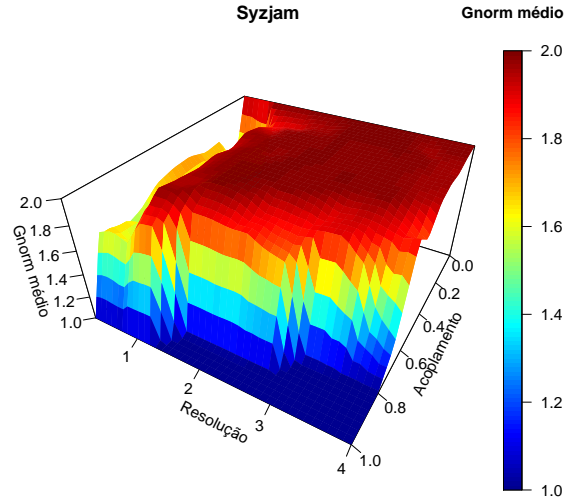


(d) Vribit .  $G_{norm} = 0.978$

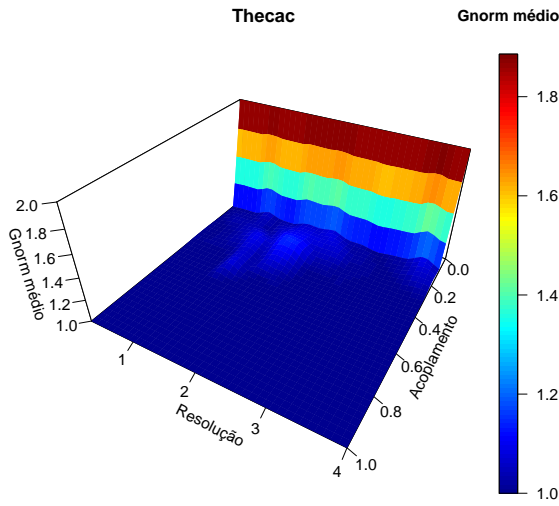
**Figura 9:** Exemplos de curvas do decaimento de  $\bar{G}$  em relação a  $\omega$  e  $\gamma$  para diferentes valores de  $\gamma$  da rede Morcegos-plantas . (a) Curvas de  $\bar{G}$  da espécie com maior valor de  $G_{norm}$  da rede. (b) Segundo maior valor de  $G_{norm}$ . (c) Valor de  $G_{norm}$  mais proximo da média geral da rede. (d) Curvas de  $\bar{G}$  referente a uma espécie com valor de  $G_{norm}$  abaixo da média da rede.



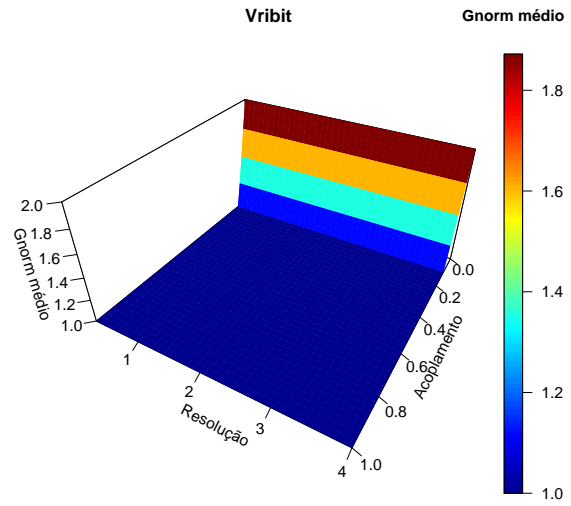
(a) Artjam .  $G_{norm} = 1.548$



(b) Syzjam .  $G_{norm} = 1.538$

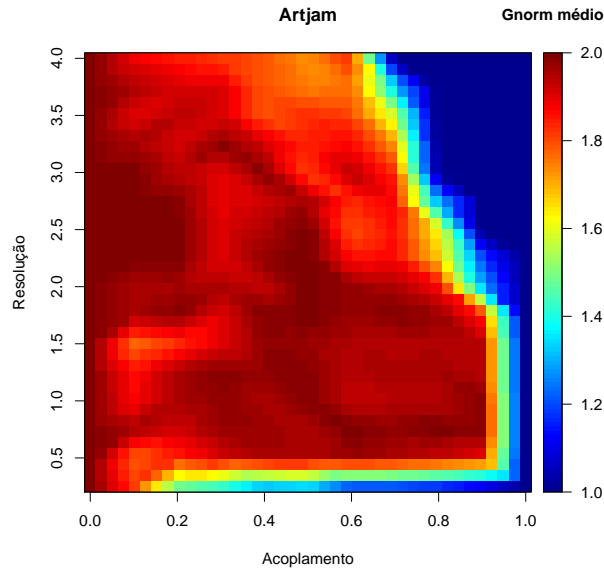


(c) Thecac .  $G_{norm} = 0.990$

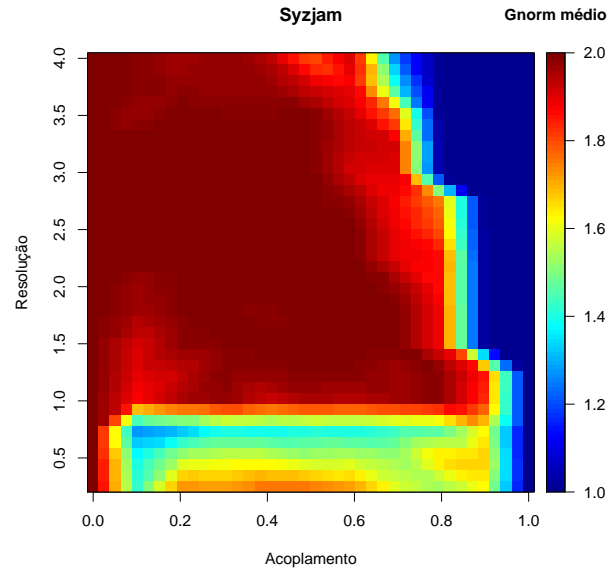


(d) Vribit .  $G_{norm} = 0.978$

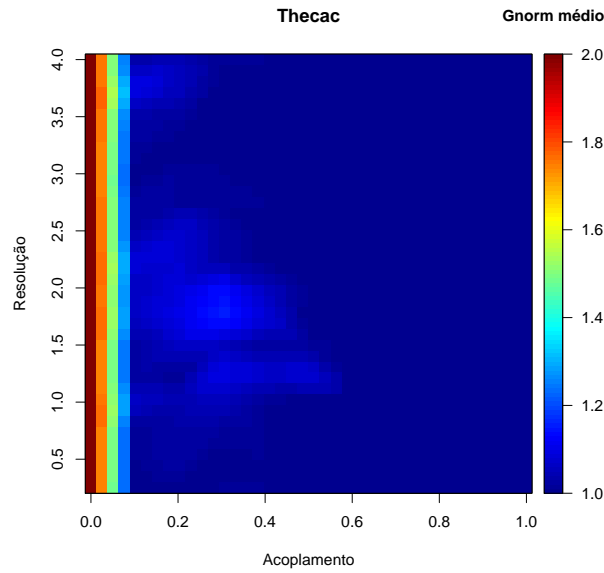
**Figura 10:** Exemplos de superfícies do decaimento de  $\bar{G}$  em relação a  $\omega$  e  $\gamma$  para diferentes valores de  $\gamma$  da rede Morcegos-plantas . (a) Superfície de  $\bar{G}$  da especie com maior valor de  $G_{norm}$  da rede. (b) Segundo maior valor de  $G_{norm}$ . (c) Valor de  $G_{norm}$  mais proximo da média geral da rede. (d) Superfície de  $\bar{G}$  referente a uma espécie com valor de  $G_{norm}$  abaixo da média da rede.



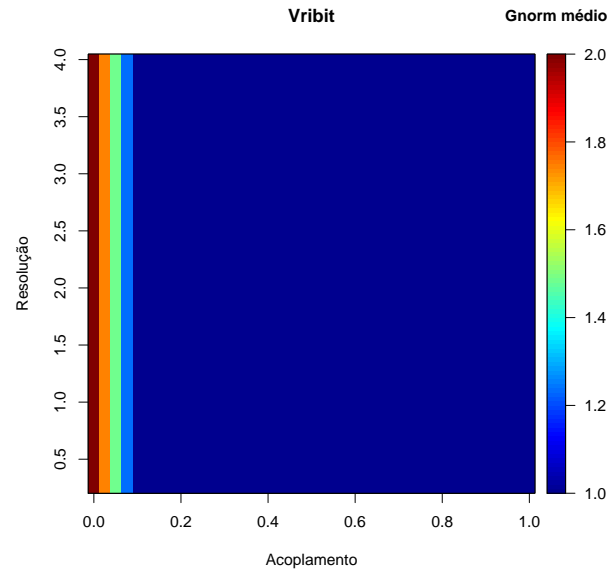
(a) Artjam .  $G_{norm} = 1.548$



(b) Syzjam .  $G_{norm} = 1.538$



(c) Thecac .  $G_{norm} = 0.990$



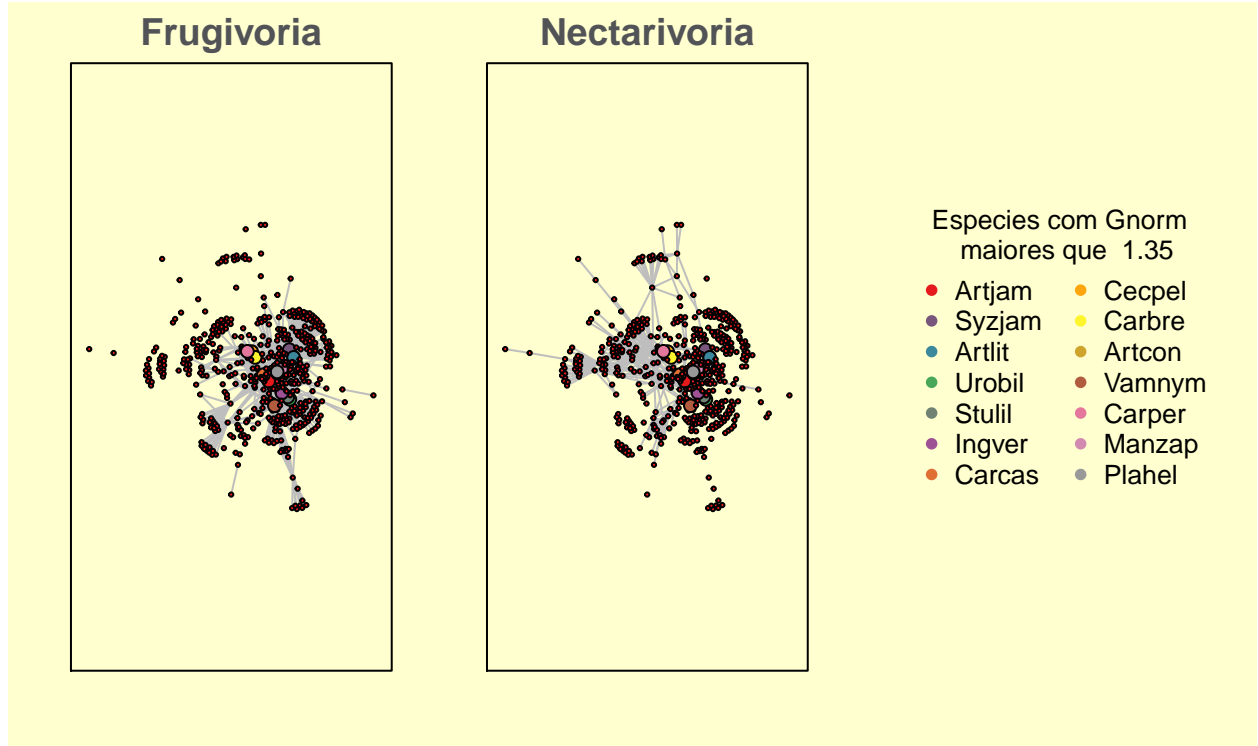
(d) Vribit .  $G_{norm} = 0.978$

**Figura 11:** Exemplos de mapas de calor do decaimento de  $\bar{G}$  em relação a  $\omega$  e  $\gamma$  para diferentes valores de  $\gamma$  da rede Morcegos-plantas . (a) Mapa de calor de  $\bar{G}$  da espécie com maior valor de  $G_{norm}$  da rede. (b) Segundo maior valor de  $G_{norm}$ . (c) Valor de  $G_{norm}$  mais proximo da média geral da rede. (d) Mapa de calor de  $\bar{G}$  referente a uma espécie com valor de  $G_{norm}$  abaixo da média da rede.



### 8.3 Seleção das espécies com maior $G_{norm}$ .

A figura 12 e a tabela 4 mostram as espécies com valor de  $G_{norm}$  acima de 1.35, ou seja, aquelas com decaimento de  $G$  mais lento da rede Morcegos-plantas.



**Figura 12:** Espécies com  $G_{norm}$  maiores que 1.35 em destaque de tamanho e cor.

**Tabela 4:** Espécies com valores de  $G_{norm}$  maiores que 1.35

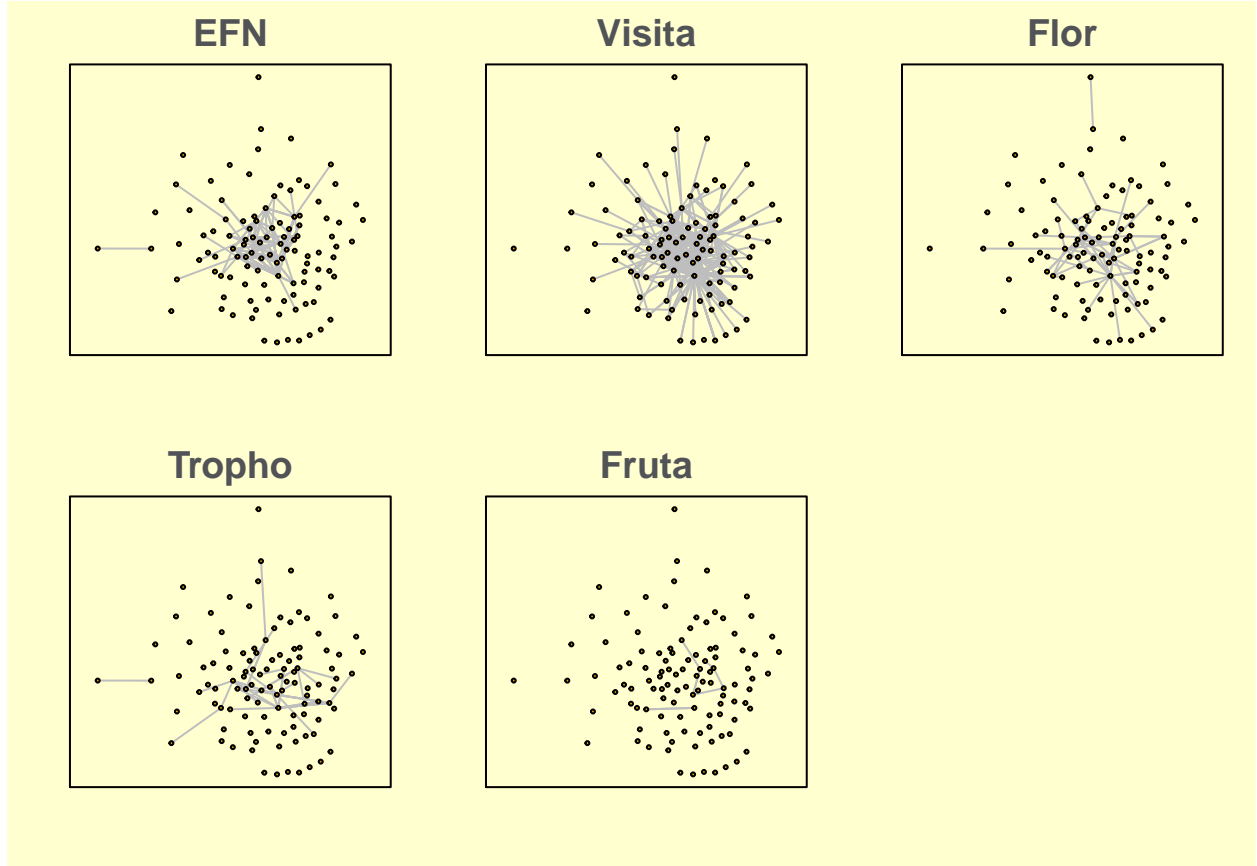
Espécie	$G_{norm}$
Artjam	1.548
Syzjam	1.538
Artlit	1.509
Urobil	1.488
Stulil	1.469
Ingver	1.432
Carcas	1.401
Cecpel	1.401
Carbre	1.392
Artcon	1.391
Vamnym	1.365
Carper	1.364
Manzap	1.360
Plahel	1.359

## 9 Visão geral da rede Formigas-planta

A rede Formigas-planta (Costa et al., 2016) possui 5 camadas (Fruta , Tropho , Flor , EFN e Visita), 108 nós e 370 conexões. A tabela 5 mostra o resumo das propriedades da rede. A figura 13 apresenta uma visão geral da rede.

**Tabela 5:** Propriedades da rede Formigas-planta

Propriedade	Valor
Número de Camadas	5
Tipo de conexões	Fruta , Tropho , Flor , EFN e Visita
Número de nós	108
Número de conexões	370



**Figura 13:** Visão geral da rede Formigas-planta .

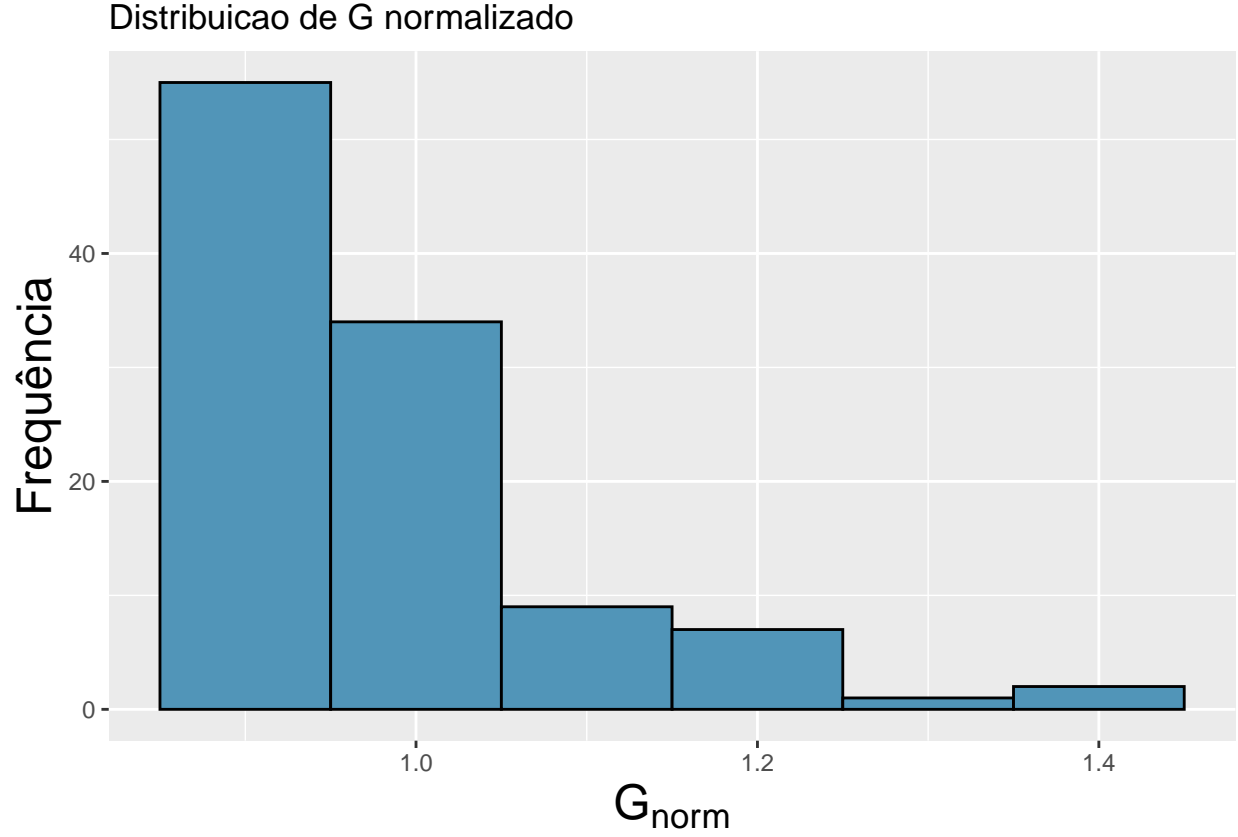
## 10 Resultados da rede Formigas-planta

### 10.1 Distribuição de $G_{norm}$

A variável  $G$  foi calculada para 10 partições de  $\omega$ , ou seja, o tamanho do passo dado dentro de  $\omega$  foi de 0.1. O processo foi repetido para 16 partições de  $\gamma$ , com  $\gamma$  começando em 0.25, com passos de -0.25 até um  $\gamma$  máximo de 4. O cálculo de  $\bar{G}$  foi feito usando 100 iterações. A tabela 6 resume os parâmetros de execução do código e a figura 14 mostra a distribuição dos valores de  $G_{norm}$  médio obtidos.

**Tabela 6:** Parâmetros de execucao

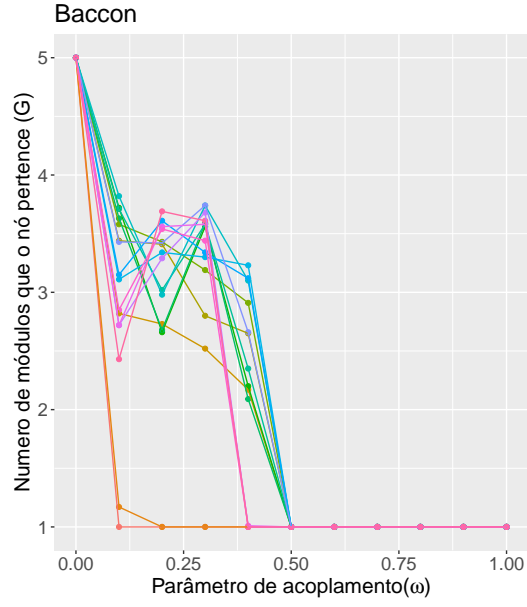
Parâmetro	Valor
Iterações	100
Partições de omega	10



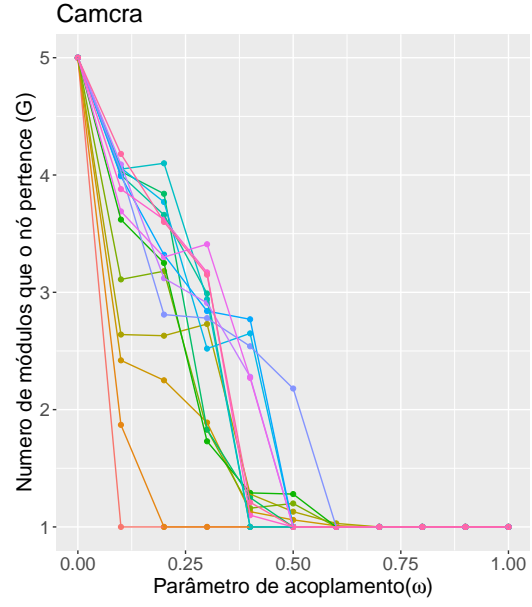
**Figura 14:** Distribuição de  $G_{norm}$  médio da rede Formigas-planta .

## 10.2 Variação de $\overline{G}$ por $\omega$

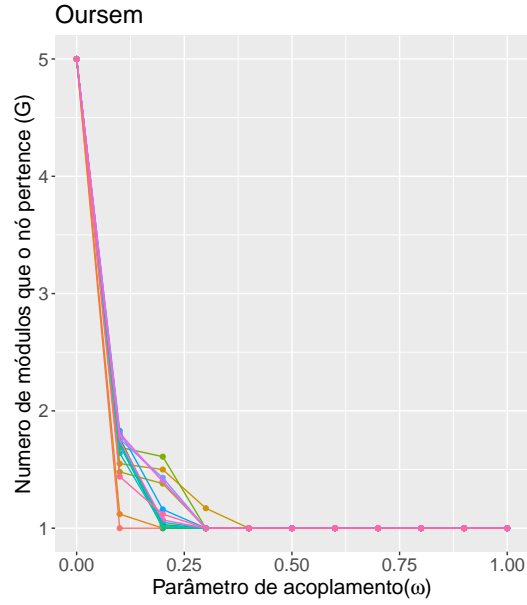
Como temos dados em 3 dimensões  $(\overline{G}, \omega, \gamma)$  temos algumas formas diferentes para apresentar os valores de  $\overline{G}$  em relação a  $\omega$  e  $\gamma$ , não sei dizer se devemos usar uma delas, as três ou alguma outra. A figura 15 mostra curvas de decaimento de  $\overline{G}$  por  $\omega$  para diferentes nós com diferentes valores de  $G_{norm}$  e para diferentes valores de  $\gamma$ . A figura 16 mostra a superfície 3D formada por  $\overline{G}$  em relação a  $\omega$  e  $\gamma$ . A figura 17 mostra a mesma superfície da figura 16 mas no formato de mapa de calor.



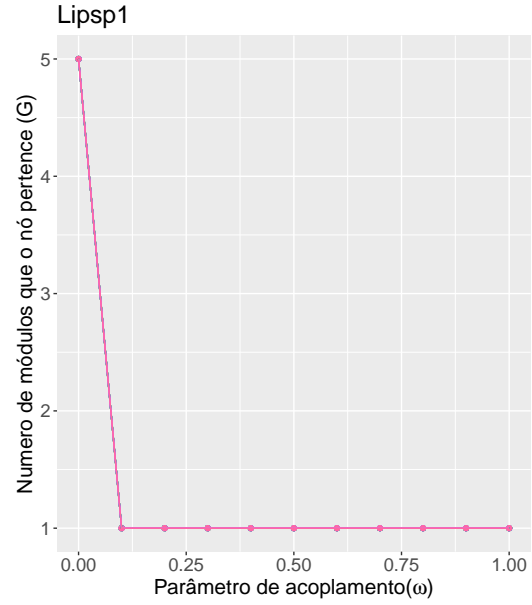
(a) Baccon .  $G_{norm} = 1.386$



(b) Camcra .  $G_{norm} = 1.355$

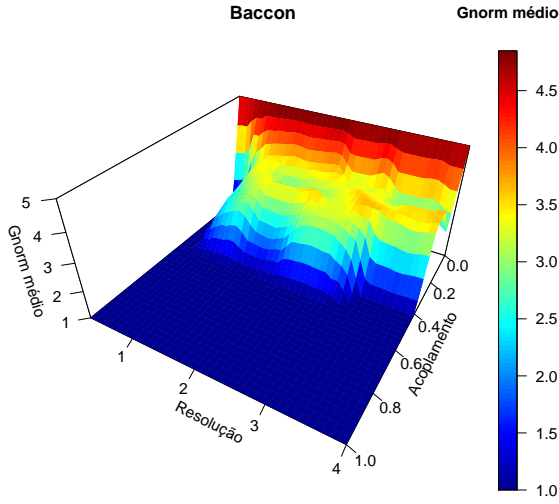


(c) Oursem .  $G_{norm} = 0.997$

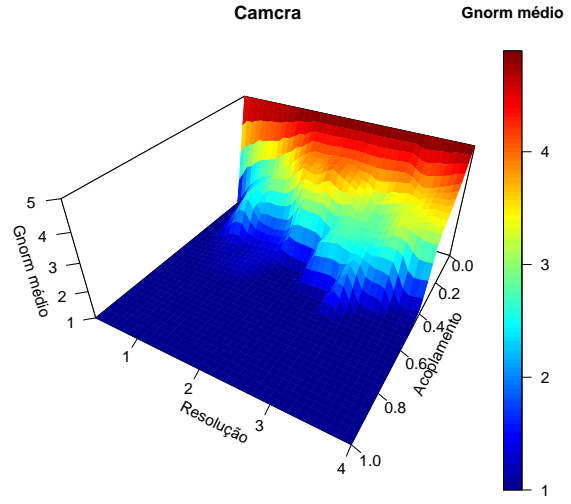


(d) Lipsp1 .  $G_{norm} = 0.945$

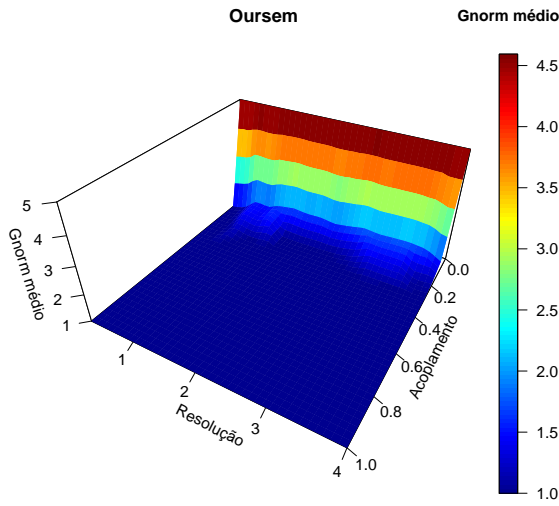
**Figura 15:** Exemplos de curvas do decaimento de  $\bar{G}$  em relação a  $\omega$  e  $\gamma$  para diferentes valores de  $\gamma$  da rede Formigas-planta . (a) Curvas de  $\bar{G}$  da espécie com maior valor de  $G_{norm}$  da rede. (b) Segundo maior valor de  $G_{norm}$ . (c) Valor de  $G_{norm}$  mais proximo da média geral da rede. (d) Curvas de  $\bar{G}$  referente a uma espécie com valor de  $G_{norm}$  abaixo da média da rede.



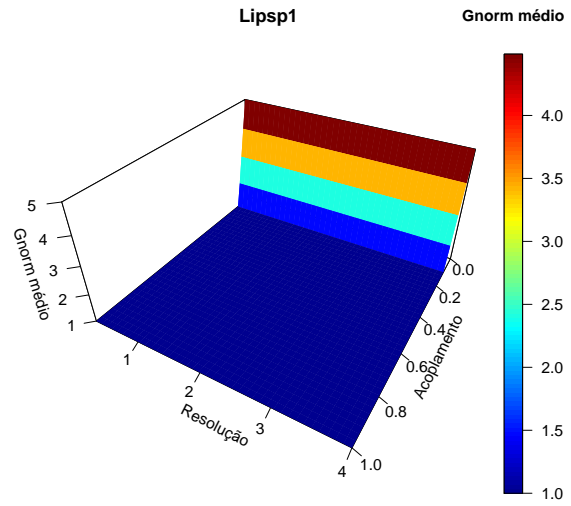
(a) Baccon .  $G_{norm} = 1.386$



(b) Camcra .  $G_{norm} = 1.355$

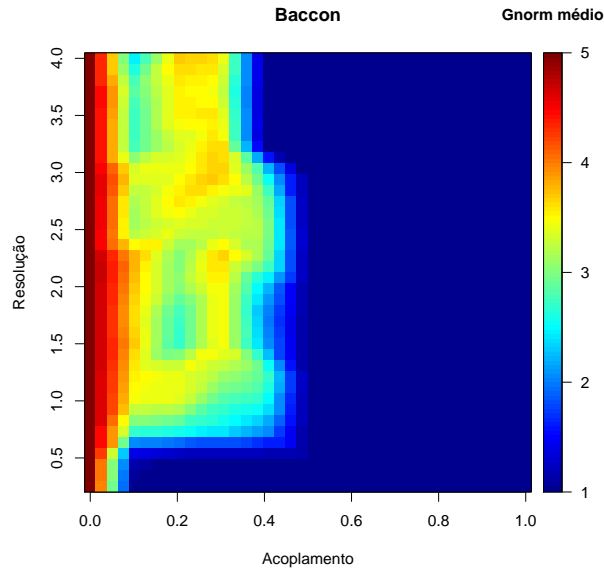


(c) Oursem .  $G_{norm} = 0.997$

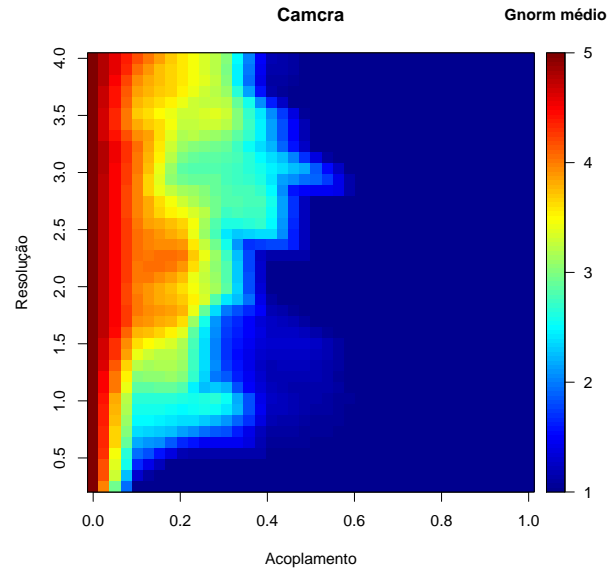


(d) Lipsp1 .  $G_{norm} = 0.945$

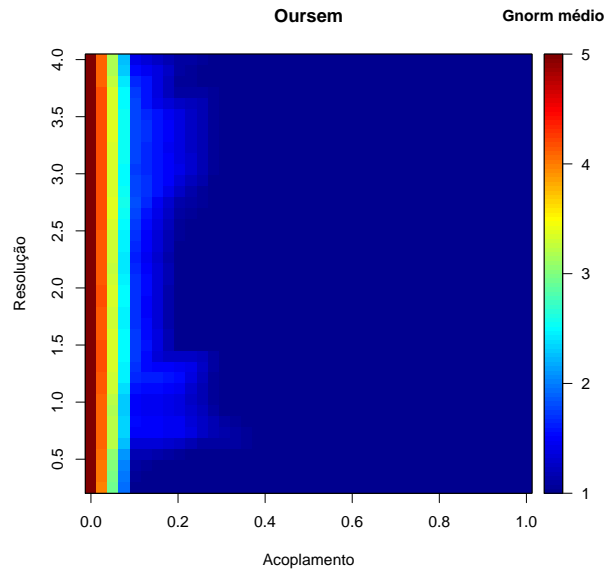
**Figura 16:** Exemplos de superfícies do decaimento de  $\bar{G}$  em relação a  $\omega$  e  $\gamma$  para diferentes valores de  $\gamma$  da rede Formigas-planta . (a) Superfície de  $\bar{G}$  da especie com maior valor de  $G_{norm}$  da rede. (b) Segundo maior valor de  $G_{norm}$ . (c) Valor de  $G_{norm}$  mais proximo da média geral da rede. (d) Superfície de  $\bar{G}$  referente a uma espécie com valor de  $G_{norm}$  abaixo da média da rede.



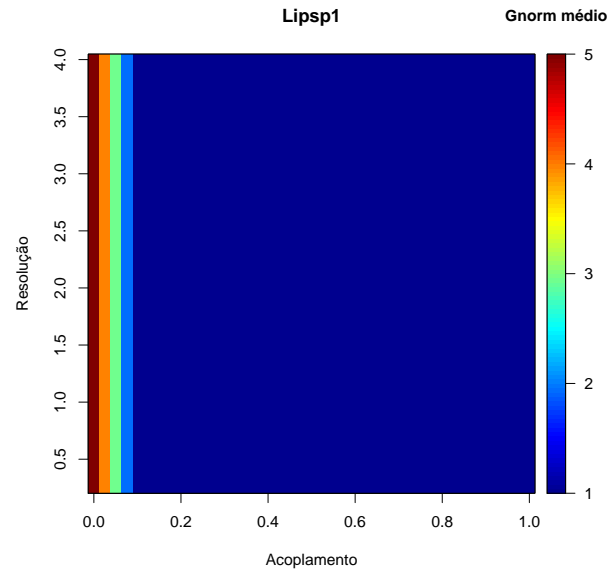
(a) Baccon .  $G_{norm} = 1.386$



(b) Camcra .  $G_{norm} = 1.355$



(c) Oursem .  $G_{norm} = 0.997$

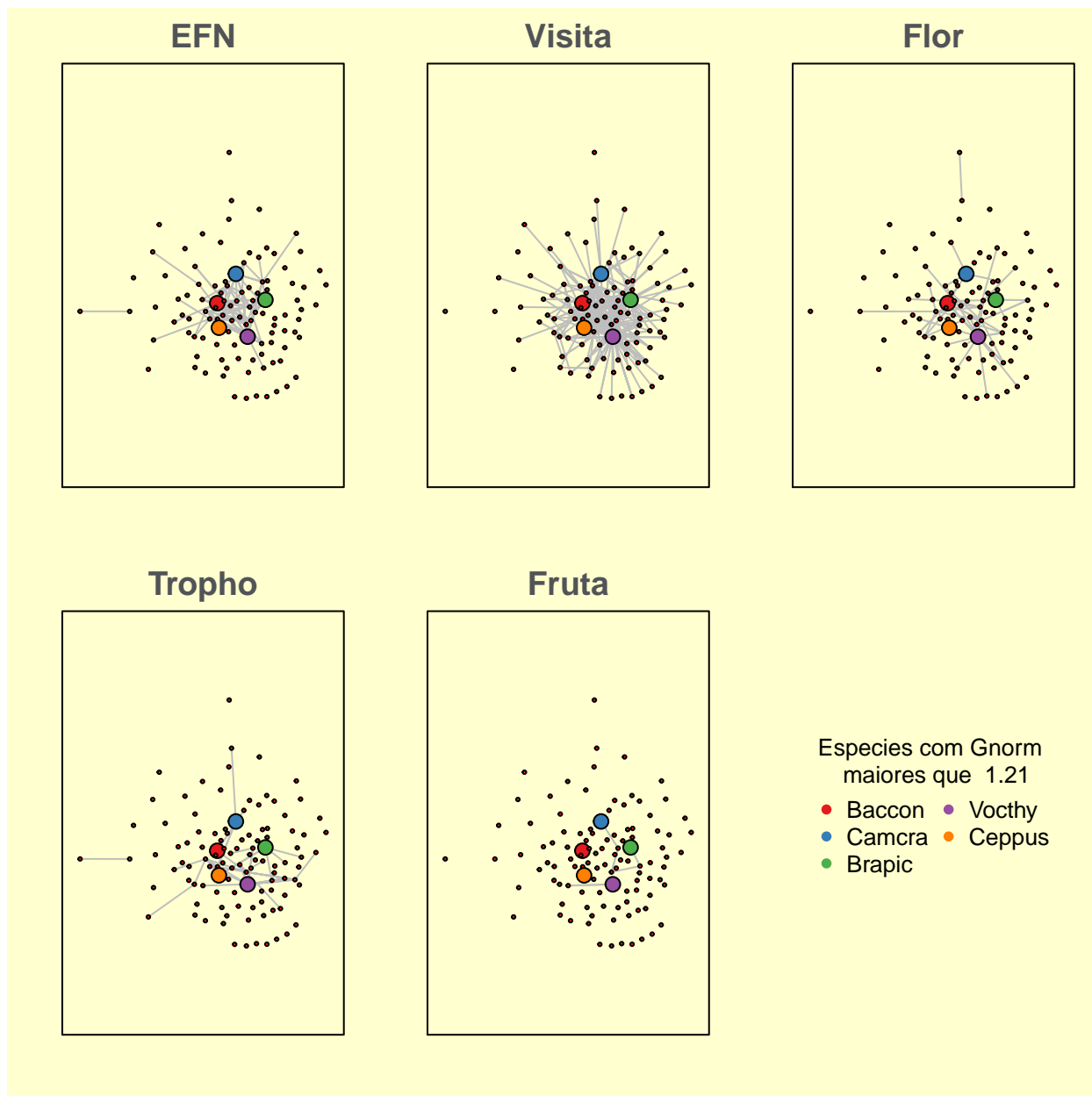


(d) Lips1 .  $G_{norm} = 0.945$

**Figura 17:** Exemplos de mapas de calor do decaimento de  $\bar{G}$  em relação a  $\omega$  e  $\gamma$  para diferentes valores de  $\gamma$  da rede Formigas-planta . (a) Mapa de calor de  $\bar{G}$  da espécie com maior valor de  $G_{norm}$  da rede. (b) Segundo maior valor de  $G_{norm}$ . (c) Valor de  $G_{norm}$  mais próximo da média geral da rede. (d) Mapa de calor de  $\bar{G}$  referente a uma espécie com valor de  $G_{norm}$  abaixo da média da rede.

### 10.3 Seleção das espécies com maior $G_{norm}$ .

A figura 18 e a tabela 7 mostram as espécies com valor de  $G_{norm}$  acima de 1.21, ou seja, aquelas com decaimento de  $G$  mais lento da rede Formigas-planta.



**Figura 18:** Espécies com  $G_{norm}$  maiores que 1.21 em destaque de tamanho e cor.



**Tabela 7:** Espécies com valores de  $G_{norm}$  maiores que 1.21

Espécie	$G_{norm}$
Baccon	1.386
Camcra	1.355
Brpic	1.341
Vochy	1.235
Ceppus	1.222

## 11 Comparação entre nós com alto $G_{norm}$ com nós centrais da rede agregada

Os nós que possuem maiores valores de  $G_{norm}$  parecem ser nós que possuem uma importância diferenciada dentro da rede, pois, ao permanecerem em diferentes grupos em diferentes camadas, mesmo após o aumento do acoplamento, mostra que os nós possuem uma relação mais igualitária entre diferentes grupos em camadas diferentes, o que os tornam nós com potencial de bons conectores entre diferentes grupos em camadas diferentes da rede.

Uma outra métrica que busca classificar as espécies mais importantes dentro de uma rede é a centralidade [ref?]. Se compararmos os nós mais centrais de uma rede com os nós com maior  $G_{norm}$ , podemos verificar a relação entre as métricas.

Existem vários tipos de centralidade diferentes, calculadas usando diferentes parâmetros como base. Neste trabalho vamos comparar quatro delas com os nós que possuem alto  $G_{norm}$ . São elas: *Closeness* [ref?], *Betweenness* [ref?], *Eigen Vector* [ref?] e *Degree*. Estas centralidades são calculadas sobre a rede multicamada agregada mantendo os links duplos, caso existam.

Duas formas de comparação foram usadas. Uma que se trata apenas de uma comparação binária, levando em conta apenas o número de nós em comum entre as centralidades e alto  $G_{norm}$ . Outra que leva em conta a distância entre rankings (posições) dos nós selecionados entre os mais centrais e altos valores de  $G_{norm}$ .

A similaridade binária é bastante trivial e seu índice é calculado da seguinte forma:

$$Sim_{bin} = \frac{\sum j^N \sum_i^N \delta(c_j, g_i)}{N} \quad (8)$$

Onde  $\delta(c_j, g_i)$  é o delta de kronecker;  $N$  é o ranking de corte escolhido, e.g  $N = 5$  no caso de escolhermos avaliar os cinco nós melhores ranqueados de cada métrica;  $c_j$  e  $g_i$  são os vetores ordenadas por ranking dos nós mais centrais e com maiores valores de  $G_{norm}$

respectivamente. Os valores da similaridade variam entre 0 (nenhum nó em comum) e 1 (todos os nós em comum).

Além da comparação binária, podemos avaliar a similaridade levando em conta ranking do nó dentro do conjunto de corte escolhido. Para isso basta adicionarmos um fator de peso que é proporcional ao inverso da distância entre os nós dentro da equação 8.

$$Sim_{dist} = \frac{\sum_j^N \sum_i^N \frac{1}{1+|i-j|} \delta(c_j, g_i)}{N} \quad (9)$$

Onde o valor absoluto de  $i - j$  é equivalente a distância entre os rankings dos nós  $g_i$  e  $c_j$ . Os valores de  $Sim_{dist}$  variam entre 0 e 1, onde 0 corresponde a nenhum nó em comum entre os métodos e 1 corresponde a todos os nós em comum e com o mesmo ranking.

## 11.1 Comparação das centralidades com alto $G_{norm}$ para a rede Morcegos-Plantas

**Tabela 8:** Nós ordenados de forma decrescente de acordo com seus valores de centralidade e  $G_{norm}$  para a rede de Morcegos-plantas.

Closeness									
Artjam	Carper	Cecobt	Cecpel	Artlit	Mactin	Manind	Manzap	Piparb	Carpap
0.000822	0.000792	0.000787	0.000767	0.000765	0.000755	0.000746	0.000743	0.000740	0.000737
Betweeness									
Artjam	Carper	Artlit	Glosor	Stulil	Lepcur	Ceipen	Anocau	Syzjam	Stulud
0.3035	0.2642	0.1523	0.1441	0.1334	0.0803	0.0700	0.0697	0.0555	0.0552
Eigen Vector									
Artjam	Carper	Artlit	Stulil	Cecpel	Glosor	Carcas	Cecobt	Manzap	Ficins
1.0000	0.8626	0.8017	0.3961	0.3685	0.3119	0.2744	0.2690	0.2664	0.2603
Degree									
Artjam	Carper	Artlit	Stulil	Glosor	Carcas	Lepcur	Phyhas	Anocau	Carbre
149	138	112	72	65	44	43	40	30	29
Gnorm									
Artjam	Syzjam	Artlit	Urotil	Stulil	Ingver	Carcas	Cecpel	Carbre	Artcon
1.5483	1.5385	1.5092	1.4877	1.4692	1.4321	1.4014	1.4010	1.3917	1.3908

**Tabela 9:** Similaridade binária entre  $G_{norm}$  e outras métricas de centralidade.

Closeness	Betweeness	Eigen Vector	Degree
0.3	0.4	0.5	0.5

**Tabela 10:** Similaridade ponderada pela distância dos rankings entre  $G_{norm}$  e outras métricas de centralidade.

Closeness	Betweenness	Eigen Vector	Degree
0.153	0.313	0.375	0.350

## 12 Discussão preliminar

Nas duas redes reais estudadas, a distribuição dos valores de  $G_{norm}$  está como esperado, poucas espécies acima da média da rede e muitas espécies próximas da média ou abaixo dela. Isso mostra que apenas poucas espécies atingem valores altos de  $G_{norm}$ , sendo essa uma característica não comun entre nós da rede, assim como altos valores de centralidade

A variação de resolução  $\gamma$  não parece influenciar  $G$  de forma tão intensa quanto a variação no acoplamento  $\omega$ , mas ela existe e com certeza é mais robusto selecionar nós com alto  $G_{norm}$  como candidatos a centrais considerando a variação em  $\gamma$  também, pois assim minimizamos a probabilidade de estarmos observando uma exceção dentro de um grande universo de parâmetros de resolução além de obter as espécies-chave de uma forma mais generalizada.

Caso exista uma variação significativa de  $G_{norm}$  em relação a  $\gamma$ , podemos concluir que as espécies-chave detectadas por  $G_{norm}$  possuem centralidade relativa a resolução, ou seja, espécies podem ter sua centralidade aumentada ou diminuída caso estejamos interessados em grupos maiores ou menores. [Realizar testes com redes sintéticas para verificar se isso ocorre muito]

As espécies com maior  $G_{norm}$  selecionadas possuem boa intersecção com as espécies centrais descritas em Mello et al. (2019) e Costa et al. (2016). Isto é interessante, já que o ranqueamento das espécies foi feito levando em conta a resistência de cada espécie ao parâmetro de acoplamento entre camadas  $\omega$ , algo muito diferente dos métodos de centralidade tradicionais. No entanto, ainda preciso comparar numericamente valores de centralidade usando a rede agregada binária e com peso de aresta igual a quantidade de interações em cada camada.

## 13 Próximas etapas

### 13.1 Outubro

- Compreender melhor o parâmetro de acoplamento  $\gamma$  com visualização.
- Buscar uma boa forma de visualizar os módulos para rastrear a situação de um nó. Acredito que assim será possível entender melhor como estão posicionados os nós de maior  $G_{norm}$  em relação aos módulos ao variarmos  $\omega$  e  $\gamma$ . Acredito que isso ajudará no insight do que significa na prática essa variável  $G_{norm}$  alta que apenas alguns nós possuem.

### 13.2 Novembro

- Validar ou desvalidar  $G_{norm}$  como forma de localizar espécies que são boas conectoras entre camadas ou centrais de uma forma geral usando diversas redes sintéticas (usar simulação do Rafael Pinheiro) para verificar se o comportamento de  $G_{norm}$  é consistente, e compará-lo como outras métricas.
- Começar a escrita oficial do TCC.

## 14 Referências

- Bianconi, G. (2018). *Multilayer networks: Structure and function*. Oxford University Press.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/p10008>
- Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., & Wagner, D. (2008). On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2), 172–188. <https://doi.org/10.1109/tkde.2007.190689>
- Costa, F. V., Mello, M. A. R., Bronstein, J. L., Guerra, T. J., Muylaert, R. L., Leite, A. C., & Neves, F. S. (2016). Few ant species play a central role linking different plant resources in a network in rupestrian grasslands. *PLOS ONE*, 11(12), e0167161. <https://doi.org/10.1371/journal.pone.0167161>
- Good, B. H., Montjoye, Y.-A. de, & Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Physical Review E*, 81(4). <https://doi.org/10.1103/physreve.81.046106>
- Ings, T. C., & Hawes, J. E. (2018). The history of ecological networks. In *Ecological networks in the tropics* (pp. 15–28). Springer International Publishing. [https://doi.org/10.1007/978-3-319-68228-0\\_2](https://doi.org/10.1007/978-3-319-68228-0_2)
- Kent, D. V. (1978). *The rise of the medici : Faction in florence 1426-1434*. Oxford university press. <http://lib.ugent.be/catalog/rug01:000703415>
- Mello, M. A. R., Felix, G. M., Pinheiro, R. B. P., Muylaert, R. L., Geiselman, C., Santana, S. E., Tschapka, M., Lotfi, N., Rodrigues, F. A., & Stevens, R. D. (2019). Insights into the assembly rules of a continent-wide multilayer network. *Nature Ecology & Evolution*, 3(11), 1525–1532. <https://doi.org/10.1038/s41559-019-1002-3>
- Meunier, D. (2009). Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3. <https://doi.org/10.3389/neuro.11.037.2009>
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980), 876–878. <https://doi.org/10.1126/science.1184819>
- Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, 70(5).

<https://doi.org/10.1103/physreve.70.056131>

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2). <https://doi.org/10.1103/physreve.69.026113>

Pilosof, S., Porter, M. A., Pascual, M., & Kéfi, S. (2017). The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1(4). <https://doi.org/10.1038/s41559-017-0101>

Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, 74(1). <https://doi.org/10.1103/physreve.74.016110>

Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on ISI category classification. *Journal of Informetrics*, 4(2), 185–193. <https://doi.org/10.1016/j.joi.2009.11.005>