

# Recuperação Inteligente De Informações

## Relatório Final - Impacto da extinção das espécies

### Equipe Data Fund for Nature

Alexandre Herrero Matias      Lucas da Silva Nolasco  
Nicolas Abril  
Departamento Acadêmico de Informática,  
Universidade Tecnológica Federal do Paraná (UTFPR)

## 1 Introdução

A interação dos seres humanos com a natureza no período mais recente da história ajudou a contribuir para o desgaste dos recursos naturais e depredação de ambientes silvestres. Isso, por sua vez, tem apresentado diversas consequências, indo desde as alterações climáticas ao redor do globo até a extinção de algumas espécies. No entanto, apesar desse impacto já ser muito prejudicial, ele pode ainda ser potencializado. Isso acontece pelas interações que essas espécies têm em seu habitat natural, de forma que a extinção de um determinado grupo pode desequilibrar o ambiente onde eles comumente seriam encontrados e, consequentemente, também trazer prejuízos para outras espécies. Como existe uma grande quantidade de animais em risco após todos esses anos de uma relação predatória dos seres humanos com a natureza, seria interessante descobrir quais as espécies que podem causar o maior impacto à natureza caso seja extinta, para dessa forma descobrir a melhor maneira de concentrar os esforços para a preservação do meio ambiente.

Dentro desse cenário, esse trabalho tem como principal objetivo avaliar os impactos da extinção das espécies para assim identificar as mais vulneráveis. Para conduzir essa análise, foram definidas algumas perguntas de pesquisa responsáveis por nortear os rumos da pesquisa, sendo elas:

1. Uma espécie estar em risco de extinção aumenta o risco das espécies que interagem diretamente com ela?

2. É possível medir ou estimar o impacto que a extinção de uma espécie causará nos ambientes em que ela habita?
3. Existe algum padrão ou tendência que levam grupos de espécies a serem mais vulneráveis?
4. É possível medir os impactos que espécies invasoras tem nos ecossistemas em que elas são introduzidas?

A partir dessas questões e utilizando as informações disponibilizadas pelas bases Red List da IUCN, as interações fornecidas pela Global Biotic Interactions e as base de espécies invasoras GISD, foi possível chegar a algumas conclusões. A primeira delas é que há indícios de que espécies em extinção tem a tendência a se relacionar mais com outras espécies em extinção. Também foi possível observar um padrão nas relações entre as espécies, mostrando que há restrições de alimentação com base no habitat de uma espécie (marinho ou terrestre). Por fim, também foi possível identificar que a importância da espécie *Castanea dentata*, também conhecido como castanheiro-americano, para as espécies ao seu entorno, assim como o impacto que a *Batrachochytrium dendrobatidis*, uma espécie de fungo, causou em espécies de anfíbios ao invadir um ecossistema do qual não fazia parte.

## 2 Processamento de dados

### 2.1 Dados

As bases de dados selecionadas foram a IUCN Red List: responsável por fornecer os

dados da situação da espécie em termos de extinção; e a Global Biotic Interactions: fornece os dados das interações entre os diferentes grupos de seres vivos; e a GISD IUCN: responsável por informar as espécies exóticas e invasoras que impactam negativamente a biodiversidade.

## 2.2 Processamentos aplicados

### 2.2.1 Pergunta 1

A abordagem usada para essa pergunta foi descobrir se as espécies em risco de extinção interagem mais com outras espécies que também estão em risco. Para isso, foi selecionada a partição do grafo de interações contendo apenas espécies que estão na IUCN Red List. Também foi escolhida apenas a maior componente conexa, que continha 99% das interações, eliminando espécies que não se têm muita informação. Para manipular o grafo foi usada a biblioteca de python NetworkX. Como métrica para responder a pergunta, foi usada, para cada nó, a proporção de nós ameaçados dentro de uma partição que contém o nó alvo divididos pela proporção de nós ameaçados no grafo como um todo. Essa métrica, que foi chamada de "Prevalência de interações com espécies ameaçadas". Alterando a partição usada para calcular a métrica é possível capturar diferentes estruturas dentro do grafo. Quando a partição é baseada na vizinhanças, ela representa quantas vezes acima do esperado o vértice interage com outros vértices ameaçados, assumindo que as interações são distribuídas aleatoriamente. Para a análise dessa pergunta, as partições usadas foram os vértices com distância até 1 (vizinhança), distância até 2, até 3 e os vértices dentro da comunidade. Para calcular as comunidades foi usado um algoritmo de propagação de labels disponível no NetworkX.

Também foi feita uma análise visual, gerando diagramas com o Gephi, como forma de confirmar e validar os resultados numéricos mostrados pela métrica proposta.

### 2.2.2 Pergunta 2

Para responder esta pergunta, foram utilizados três diferentes métricas que auxiliaram no entendimento de como uma espécie impacta o ambiente em que vive. O primeiro passo

para a análise foi a criação da rede de interações ecológicas das espécies, o grafo resultante do cruzamento dos dados da IUCN Red List com os da Global Biotic Interactions. Um problema já foi encontrado neste ponto visto que o grafo obtido era extremamente denso e de difícil visualização. Para contornar este problema foi feito um corte no grafo onde foram selecionados apenas nós que representavam a relação ecológica de alimentação e que continham pelo menos uma espécie em risco crítico de extinção.

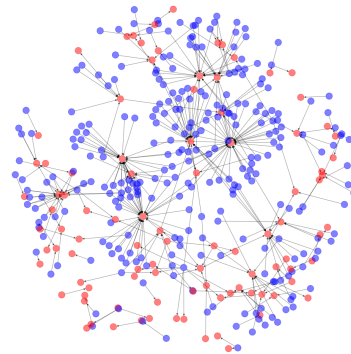


Figura 1: Grafo de interações obtido. Nós vermelhos representam espécies em risco crítico de extinção

Em seguida foram aplicadas as seguintes métricas em cada um dos nós do grafo:

1. Métrica de "Vulnerabilidade" do nó. Esta métrica foi apelidada de "Vulnerabilidade" por nós pois seu cálculo se faz removendo cada um dos nós do grafo e calculando a eficiência global do grafo após sua remoção.
2. Métrica de centralidade de grau. É uma métrica bem conhecida onde o seu valor é dado pelo número de arestas que chegam em um determinado nó (para grafos direcionados).
3. Métrica de centralidade de Katz. É outra métrica de centralidade, porém está leva em consideração o número de arestas de todos os nós que estão sendo ligados a um nó.

### 2.2.3 Pergunta 3

Buscando identificar padrões entre as espécies, foram selecionadas as interações de

alimentação entre as espécies presentes na base de interações. O objetivo com isso foi diminuir o tamanho do grafo geral de interações mantendo as relações mais importantes. Nesse grafo resultante, foi aplicado o algoritmo de clusterização Leiden principalmente pela sua característica de possuir uma menor complexidade computacional. O resultado obtido foram dois agrupamentos principais com alguns grupos menores ao redor, como mostra a Figura 8.

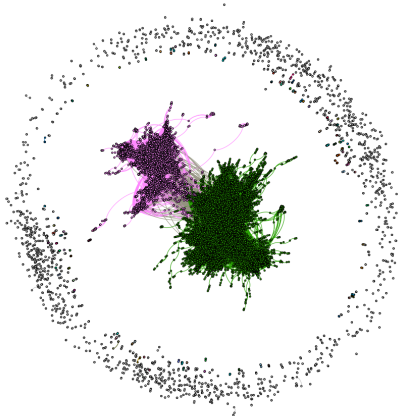


Figura 2: Agrupamentos encontrados nas interações de alimentação

Em seguida, foram analisadas algumas métricas com o objetivo de identificar a influência das espécies vulneráveis dentro desses grupos maiores e também ajudar a caracterizar esses agrupamentos. Entre as métricas calculadas que apresentaram os melhores resultados estão o prestígio do grau de entrada, que mede a quantidade de espécies para qual uma espécie de alimento, e o prestígio de domínio, que mede a quantidade de espécies que se alimentam de uma espécie e assim por diante.

Outras métricas que apresentaram resultados interessantes foram a Hubs e a Authority. Ambas fazem parte do HITS, uma métrica criada originalmente para avaliar o desempenho de páginas web. O Authority visa medir a qualidade do conteúdo existente em uma página e, dentro do contexto analisado, pode ser interpretado como a importância de uma espécie dentro da rede de interações. Já o Hubs foi criado para medir a importância de um link saindo de uma página web para outra e, por conta disso, pode ser interpretado como a importância de uma espécie se alimentar de outras.

#### 2.2.4 Pergunta 4

Inicialmente a ideia era verificar a possibilidade de identificar espécies invasoras apenas utilizando as informações de localização geográfica dos dados da IUCN Red List. Porém esta hipótese foi rapidamente descartada pois ao fazer o cruzamento destes dados com os dados da GISD IUCN (para validação) foi percebido que as poucas espécies em comum existentes estavam sem nenhuma informação de localização. Desta forma a terceira base de dados IUCN GISD teve de ser adicionada ao projeto a fim de se realizar análises sobre espécies invasoras.

Para realizar as análises as mesmas três métricas apresentadas na pergunta 2 foram utilizadas novamente (de forma invertida visto que os nós de interesse são as origens das interações) para identificar a influencia de uma espécie invasora no seu ecossistema. O primeiro passo foi cruzar os dados da GISD IUCN com os dados da Global Biotic Interactions de forma que as espécies que realizam as interações seja sempre espécies invasoras, em seguida também foram cruzados os dados da IUCN Red List de forma que as espécies que sofrem as interações estejam ameaçadas de alguma forma. o grafo de interações resultante era muito denso para que pudesse se analisado de maneira satisfatória com as métricas desejadas, desta forma também foi feito um corte onde foram selecionadas apenas interações de alimentação.

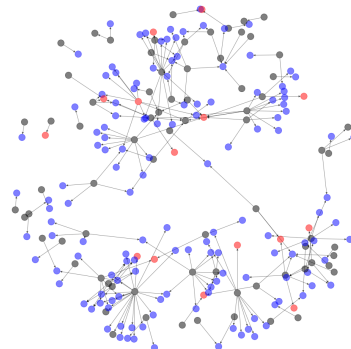


Figura 3: Grafo de interações obtido. Nós cinza representam espécies invasoras

## 3 Resultados

### 3.1 Pergunta 1

Fazendo a análise visual, percebe-se que as espécies ameaçadas estão mais ou menos agrupadas em alguns grupos, apesar de não haver uma separação tão forte, como mostra a Figura 4. É notável que algumas áreas são mais azuis, enquanto outras são mais vermelhas.

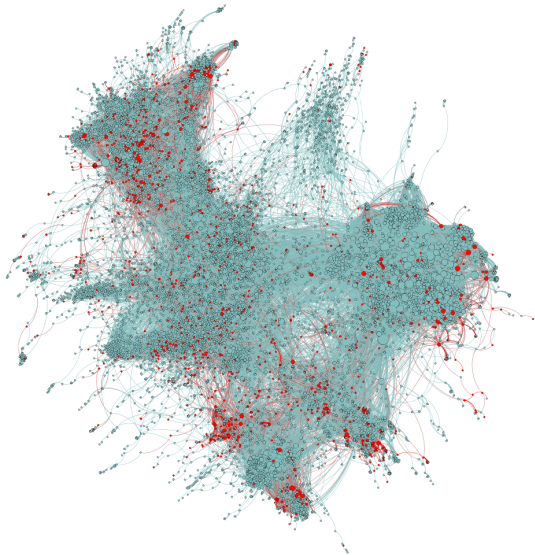


Figura 4: Grafo de interações entre espécies do IUCN Red List. Os pontos vermelhos representam espécies em risco de extinção.

A métrica de "prevalência de ameaçados" foi calculada para cada vértice do grafo usando os 4 tipos de partição descritos. Esses valores foram agregados em espécies ameaçadas e não-ameaçadas. Se as espécies ameaçadas realmente interagem mais com outras ameaçadas, a prevalência dentro da sua vizinhança será maior que 1 e se as espécies ameaçadas formam grupos com maior quantidade de ameaçados, a métrica para as outras partições terá prevalência maior que 1. A hipótese de que esses efeitos ocorreriam se comprovou verdadeira, como mostra a Tabela 1. Dentre as quatro partições, a que teve a maior prevalência de espécies ameaçadas, portanto melhor representando os agrupamentos de espécies ameaçadas, foi os vértices com distância até 2, com 78% mais interações com espécies ameaçadas do que o esperado. Curiosamente, a prevalência dos vizinhos entre as espécies não-ameaçadas também foi maior que

1, diferente do obtido para as outras partições.

Tabela 1: Valor médio da prevalência de espécies ameaçadas

Ameaçadas	$d \leq 1$	$d \leq 2$	$d \leq 3$	Comunidade
Sim	1,71	1,78	1,07	1,61
Não	1,07	0,79	0,96	0,92

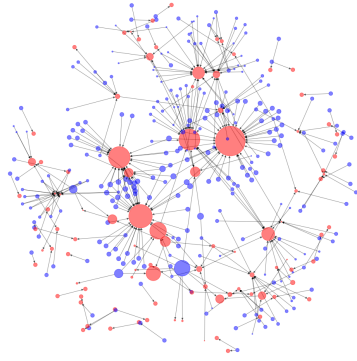
### 3.2 Pergunta 2

Aplicando os resultados obtidos em cada uma das métricas como fator de peso na rede de interações é possível visualizar quais espécies se mostram as mais influentes na rede. Para cada uma das métricas aplicadas os resultados foram relativamente similares, em todas as métricas as três espécies mais influentes foram *Castanea dentata*, *Fraxinus americana* e *Fraxinus pennsylvanica* nesta ordem.

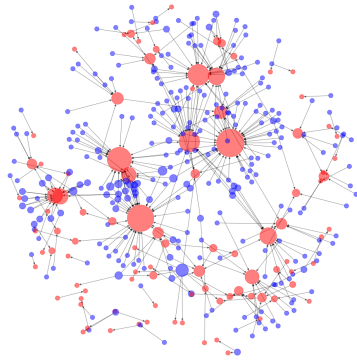
As três espécies mais influentes em todos os resultados apresentaram uma característica em comum que pode nos indicar um padrão que determina a influência de uma espécie, as três eram do reino *plantae*, ou seja plantas. Isto faz sentido quando pensamos em cadeia alimentar, onde normalmente as plantas são a base. Além disso outra característica em comum para estas espécies é o fato de ambas estarem em risco crítico de extinção, isto nos indica que estas seriam as espécies nas quais os esforços de preservação devem ser concentrados visto que sua extinção ira impactar um grande número de outras espécies.

### 3.3 Pergunta 3

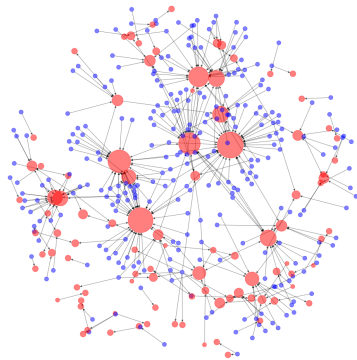
A avaliação das métricas de prestígio mostrou uma grande quantidade de plantas se destacando dentro do agrupamento verde enquanto para o agrupamento rosa apareceu uma grande quantidade de algas e outras espécies marinhas como a *Kondakovia longimana*, uma espécie de lula. Além disso, a análise da distribuição dos reinos dentro de cada grupo reforçou essa ideia, mostrando que o conjunto verde possuía uma quantidade maior de espécies dos reinos Animalia e Plantae, enquanto o conjunto rosa se destacava pela grande quantidade de espécies dos reinos Animalia e Chromista. Combinado as informações sobre os reinos que compõem cada grupo com essas espécies que obtive-



(a) Métrica de "Vulnerabilidade"



(b) Métrica de centralidade de grau



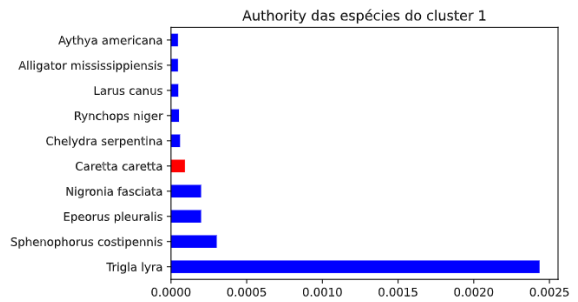
(c) Métrica de centralidade de Katz

Figura 5: Grafos resultantes da aplicação das métricas

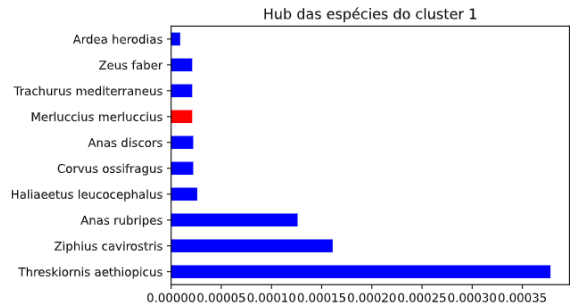
ram destaque nas métricas, foi possível concluir que o agrupamento verde representa as espécies terrestres enquanto o agrupamento rosa representa as espécies marinhas.

Já com relação às métricas Hub e Authority, elas apresentaram resultados interessantes somente entre as espécies marinhas. Como mostra a Figura 6, foi possível identi-

ficar a tartaruga marinha (*Caretta caretta*) é a espécie vulnerável de maior importância que também está em alguma situação de vulnerabilidade. Ainda foi identificado que a espécie pescada (*Merluccius merluccius*) é a espécie em vulnerabilidade que mais se destaca quando o assunto é a importância de interação com outras espécies. Ter espécies importantes com esse grau de risco pode indicar que as espécies marinhas estão em um risco maior, uma vez que essas espécies citadas anteriormente estão mais próximas da extinção e representam um papel de destaque dentro da rede de interações, podendo impactar outras espécies caso sejam extintas.



(a) Authority



(b) Hubs

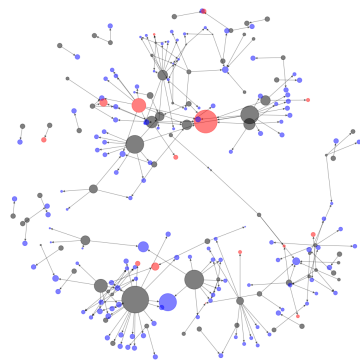
Figura 6: Hubs e Authority para as espécies marinhas. As espécies vulneráveis estão destacadas em vermelho.

### 3.4 Pergunta 4

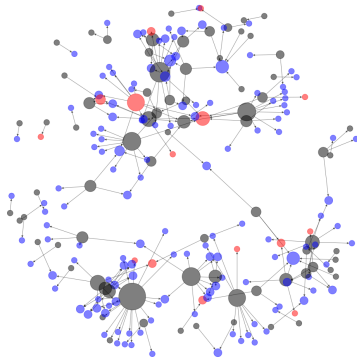
A aplicação das métricas na rede de interações nos revela que, como esperado, as espécies invasoras realmente apresentam uma grande influência nos ambiente em que são introduzidas. Nas três métricas aplicadas a espécie invasora mais influente foi *Canis lupus*, o cão doméstico, o que faz sentido quando pensamos que esta espécie foi domesticada e espalhada pelo mundo por ação do ser humano.



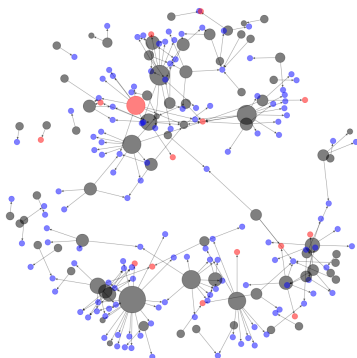
Porém algumas outras espécies invasoras também merecem destaque como o *Rattus rattus*, o rato-preto, e o *Petromyzon marinus*, a lampreia marinha ou peixe vampiro.



(a) Métrica de "Vulnerabilidade"



(b) Métrica de centralidade de grau



(c) Métrica de centralidade de Katz

Figura 7: Grafos resultantes da aplicação das métricas

Com relação a interações com espécies em risco de extinção, apenas 15% de todos os nós do grafo (sem o corte) apresentavam interação com espécies em risco. Desta forma foi

possível criar um novo grafo que contém apenas estas interações.

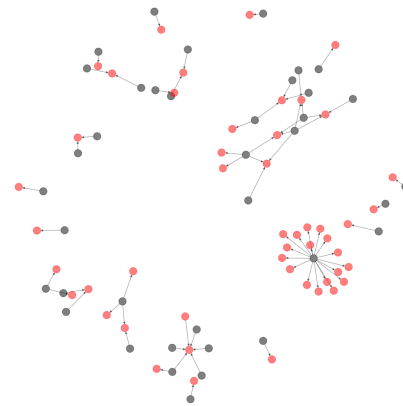


Figura 8: Grafo de interações obtido. Nós cinza representam espécies invasoras

Dentre todas estas interações as que mais chamam a atenção são as realizadas pelo *Batrachochytrium dendrobatidis* um espécie de fungo originado no leste asiático que causa uma doença conhecida como quitridiomicose em anfíbios. Estima-se que pelo menos 501 espécies de anfíbios sofrem declínios acentuados em suas populações devido ao fungo.

## 4 Limitações e trabalhos futuros

Uma das principais limitações encontradas se refere à quantidade de dados e os algoritmos usados. Aplicar os algoritmos disponíveis sobre o grafo como um todo se mostrou inviável, com operações levando na ordem de horas. Devido a isso, as análises foram feitas sobre cortes da base completa, ou usando apenas algoritmos rápidos. Isso reduziu o alcance e generalidade dos resultados. Uma direção possível para trabalhos futuros é analisar uma parcela maior dos dados e usando algoritmos mais complexos. Uma possibilidade é usar ferramentas diferentes para realizar os processamentos. Neste trabalho usamos principalmente o NetworkX e, quando possível, o Gephi, mas eles tem suas limitações.

Paradoxalmente, outro grande problema encontrado foi a base de dados das interações ser esparsa e incompleta. Como dados

dessa base tem origem em trabalhos científicos, os dados tendem a favorecer mais certos tipos de seres vivos que são mais estudados. Além disso, os dados são adicionados manualmente à base, a pedido dos organizadores ou de forma voluntária pelos pesquisadores, o que restringe seu potencial volume. Similarmente, a IUCN Red List tinha muitas espécies sem informações de localização geográfica e as espécies invasoras não tinham uma intersecção completa com a Red List. Em um trabalho futuro é possível buscar novas fontes de dados, possivelmente combinando mais de uma base para alcançar um resultado mais robusto.

Finalmente, nesse trabalho foram explorados uma quantidade relativamente pequena de métricas. É possível buscar informações novas através do uso de métricas diferentes aplicadas de formas similares. No mais, existe uma grande quantidade de outras possíveis análises a serem feitas com esses dados, como mostra a grande quantidade de trabalhos realizados sobre essas bases de dados.