

Laboratório
Nacional de
Computação
Científica



New perspectives on analyzing data from biological collections based on social network analytics

Defesa de Mestrado em Modelagem Computacional

Autor: Pedro C. de Siracusa

Orientador: Artur Ziviani

Co-orientador: Luiz M. R. Gadelha Jr.

Sumário

1. Coleções Biológicas
2. Modelos em Redes
3. Estudo de Caso
4. Conclusão e Perspectivas

Coleções Biológicas

Fontes de informação sobre biodiversidade

Smithsonian National Museum of Natural History



Dados em CBs

Espécimes físicos (vouchers) e dados contextuais de coleta

Coletados em diversas circunstâncias

Digitização de dados

Dados armazenados digitalmente

Melhor gerência e acessibilidade

Publicação de dados

Agregadores de dados, permitem acesso centralizado a dados. ex: GBIF

Exemplos de usos

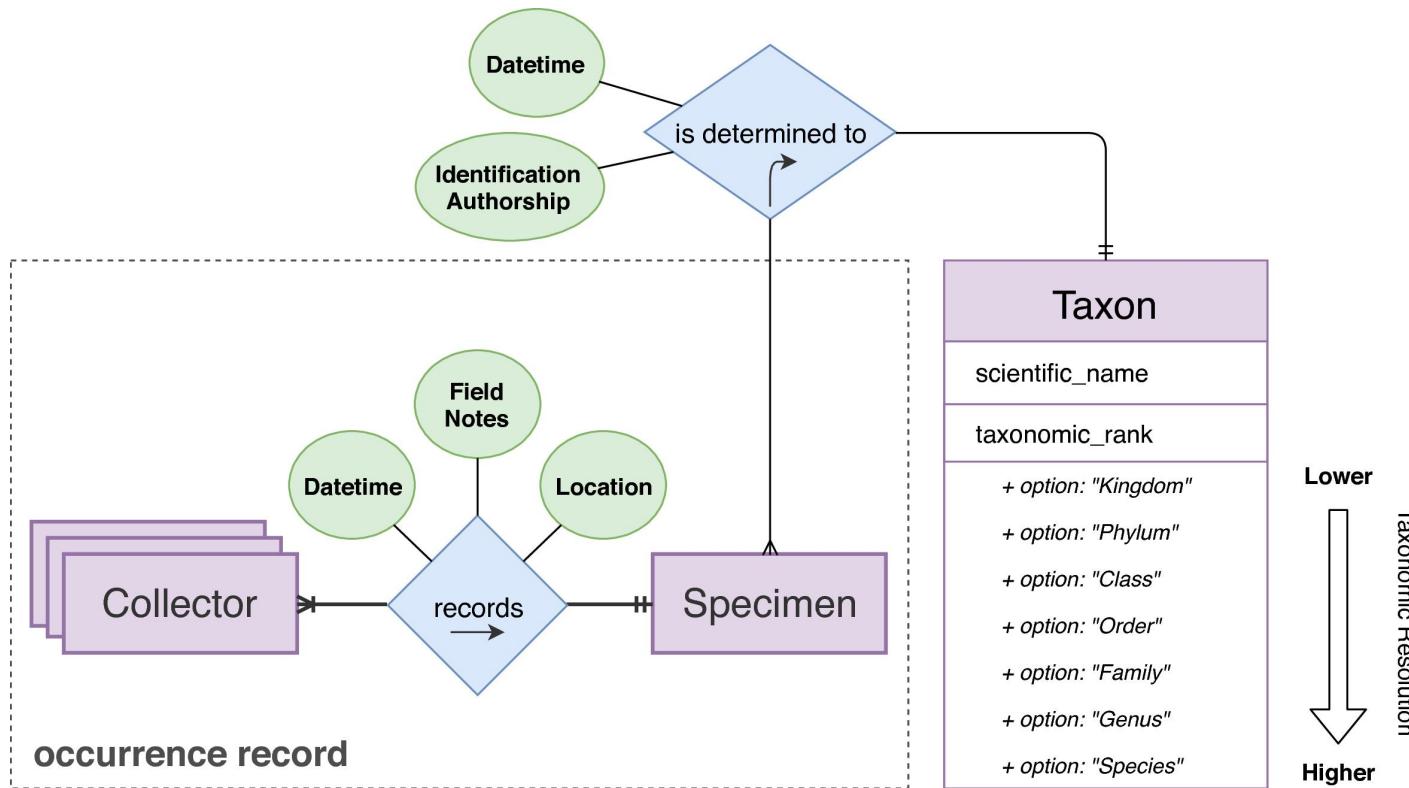
Estudar padrões de distribuição geográfica de organismos

Construir listas de espécies ameaçadas

Dados de Coleções Biológicas

Ocorrência de espécies

Armazena informação sobre o **ato de coleta**;

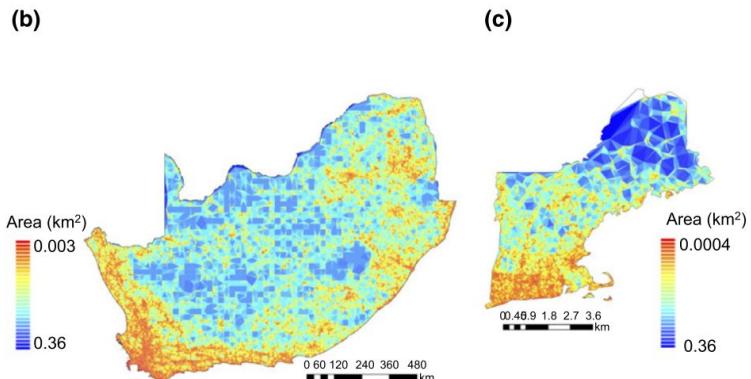
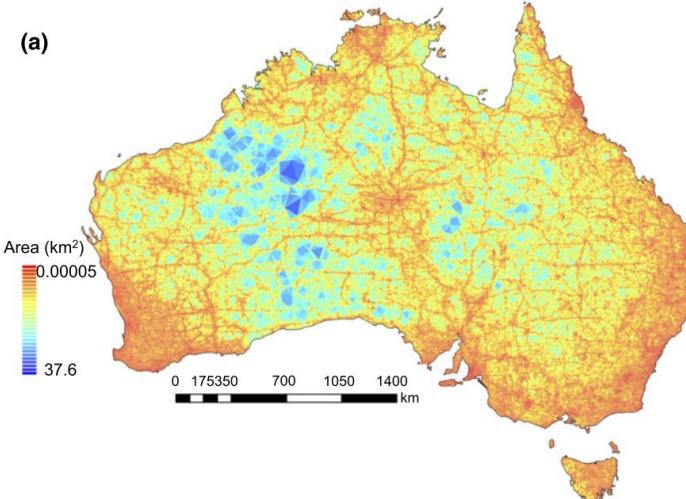


Dados de Coleções Biológicas Viés amostral

Premissas

Natureza **observacional** dos registros

Coletores possuem **interesses e comportamentos** diferenciados



Consequência

Coleções biológicas fornecem uma visão amostral **enviesada** da composição biológica existente em sua área de atuação

Variação espacial de esforço amostral em herbários na (a) Austrália, (b) África do Sul e (c) Nova Inglaterra. Retirado de [Daru et al. 2017].

Dados de Coleções Biológicas

Classes de viés

Coletor	Coletores contribuem com registros em CBs em diferentes medidas	Quais são os mais produtivos? Preferências e hábitos dos coletores são propagados para outras classes de viés
Taxonômico	Táxons não são representados nas mesmas proporções em que ocorrem em sistemas naturais	Composição taxonômica reflete os interesses dos coletores mais produtivos
Geográfico	Locais de coleta não são selecionados aleatoriamente, nem são igualmente amostrados	Locais mais amostrados são aqueles mais visitados pelos coletores mais produtivos
Temporal	Atividades de coleta não ocorrem de forma uniforme no tempo	Períodos de maior coleta tendem a ser equivalentes aos períodos de atividade dos coletores mais produtivos

Motivação e Objetivo

Motivação

Como caracterizar o viés de coletor nos dados de coleções biológicas?

Como estruturar o complexo arranjo de percepções, interesses e interações de coletores em coleções biológicas?

Objetivos

Investigar a montagem de coleções biológicas como um **processo social**

- Coletores tendem a amostrar preferencialmente espécies de seus interesses;
- Coletores interagem durante as coletas, e possivelmente se influenciam.

Modelos em Redes

Sistemas compostos por entidades e suas interações



Redes Complexas

Características topológicas complexas emergem de interações locais

Grafos

Representação matemática de redes complexas

Entidades e Interações: Vértices e Arestas

Redes Sociais

Entidades tendem a formar **comunidades**

Grafos

Modelo

$$G = (V, E)$$

V : Conjunto de vértices

E : Conjunto de arestas entre vértices

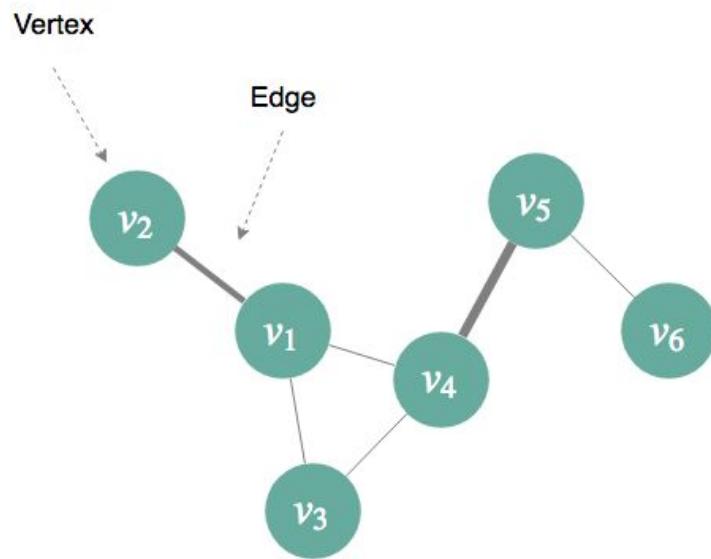
Tipos de arestas

Direcionadas: Relações são assimétricas, com um vértice-fonte e um vértice-alvo

Não-direcionadas: Relações são simétricas

Ponderadas: Relações possuem diferentes relevâncias

O **grau** (k) de um vértice é definido como o número de arestas incidentes

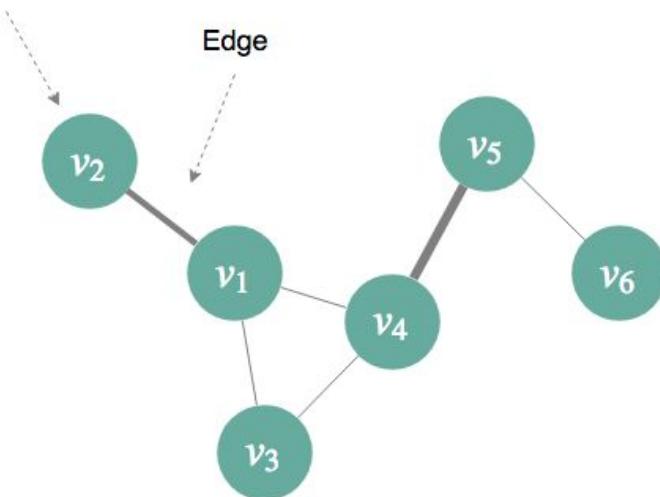


Grafos

Matriz de Adjacência

Vertex

Edge



Adjacency Matrix

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	2	1	1	0	0
v_2	2	0	0	0	0	0
v_3	1	0	0	1	0	0
v_4	1	0	1	0	3	0
v_5	0	0	0	3	0	1
v_6	0	0	0	0	1	0

Grafos

Grafos Bipartite

Modelo

$$B = (U, V, E)$$

U e V : Dois conjuntos de vértices **disjuntos** e **independentes**

E : Conjunto de arestas entre vértices

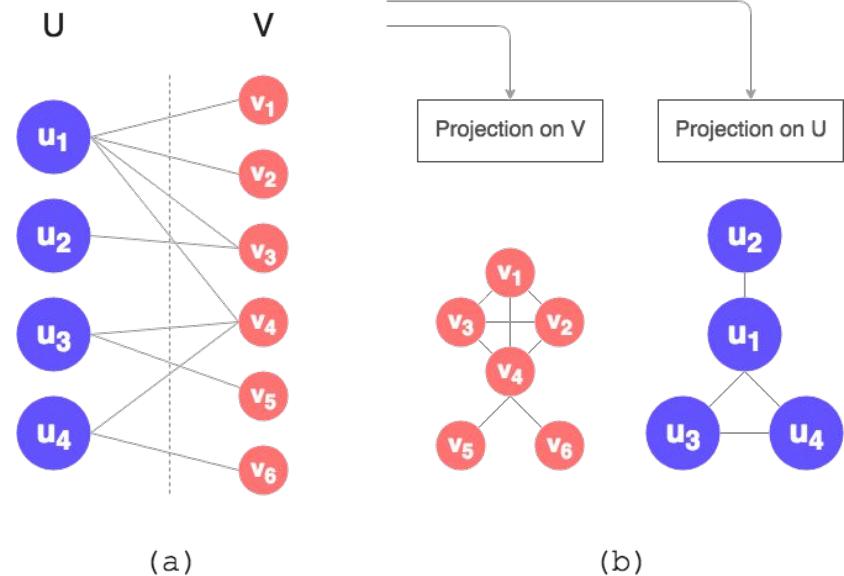
Restrição bipartite: Adjacência somente entre vértices de conjuntos opostos

Projeções bipartite

Sumarização usando apenas um dos conjuntos de vértices

Relações indiretas entre entidades

Fornecem perspectivas complementares



Nossa Contribuição

Redes Espécie-Coletor (SCN)

Modelam interesse de coletores por espécies

Redes Colaborativas de Coletores (CWN)

Modelam interações profissionais entre coletores

Modelos em Redes Redes Espécie-Coletor (SCN)

Um tipo de **redes de interesse**

Relação modelada
Coletor -[registra]-> **Espécie**

Um grafo bipartite não-direcionado

$$SCN = (S_{col}, S_{sp}, E)$$

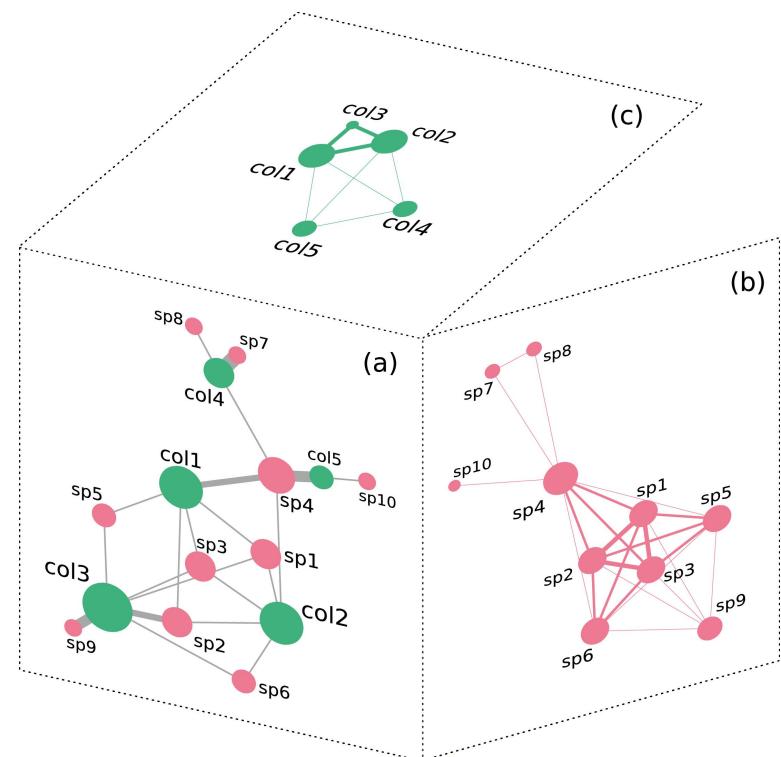
Projeções: duas perspectivas

Coletor-cêntrica

Que coletores compartilham interesse em espécies?

Espécie-cêntrica

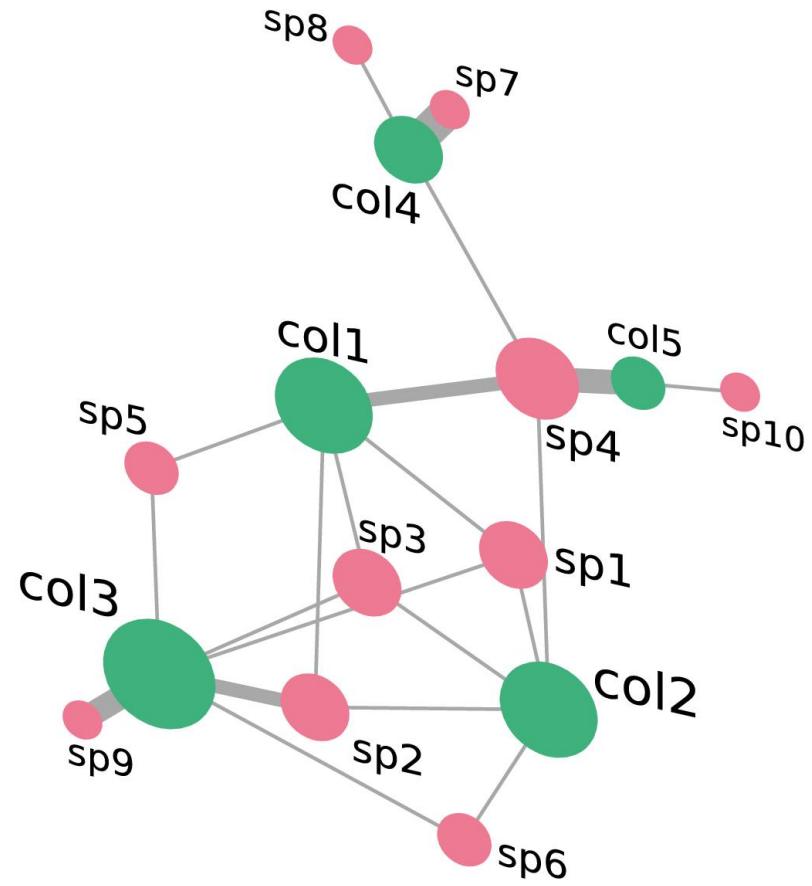
Que espécies são registradas por coletores em comum?



Aspecto geral de uma SCN. (a) Rede bipartite, com dois conjuntos de vértices. (b) Projeção sobre o conjunto de espécies S_{sp} . (c) Projeção sobre o conjunto de coletores S_{col} .

Redes Espécie-Coletor (SCN) Construção

id	recordedBy	species
0	col1; col2; col3	sp1
1	col1; col2; col3	sp2
2	col1; col2; col3	sp3
3	col1; col2	sp4
4	col1	sp4
5	col1; col3	sp5
6	col3; col2	sp6
7	col3	sp2
8	col4	sp4
9	col4	sp7
10	col4	sp7
11	col4	sp7
12	col4	sp7
13	col4	sp8
14	col3	sp9
15	col3	sp9
16	col3	sp9
17	col5	sp4
18	col5	sp4
19	col5	sp4
20	col5	sp10



Redes Espécie-Coletor (SCN)

Definições

Species Bag

A “assinatura de espécies” de um coletor

$$\sigma_{u_i} = [a_{i1}, a_{i2}, \dots, a_{im}]$$

m : número de espécies no modelo

a_{ij} : número de registros da espécie v_j pelo coletor

u_i .

	$sp1$	$sp2$	$sp3$	$sp4$	$sp5$	$sp6$	$sp7$	$sp8$	$sp9$	$sp10$
$col1$	1	1	1	2	1	0	0	0	0	0
$col2$	1	1	1	1	0	1	0	0	0	0
$col3$	1	2	1	0	1	1	0	0	3	0
$col4$	0	0	0	1	0	0	4	1	0	0
$col5$	0	0	0	3	0	0	0	0	0	1

Quorum

A “assinatura de coletores” de uma espécie

$$\iota_{v_j} = [a_{1j}, a_{2j}, \dots, a_{nj}]$$

n : número de coletores no modelo

a_{ij} : número de registros da espécie v_j pelo coletor

u_i .

Modelos em Redes

Redes Colaborativas de Coletores (CWN)

Um tipo de **redes de colaboração**

Espécies não são representadas

Relação modelada

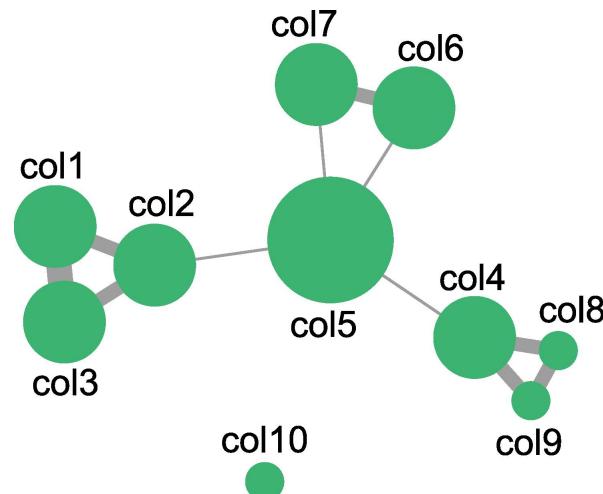
Coletor -[registra espécime com]-> **Coletor**

Um grafo não-direcionado

$$CWN = (S, E)$$

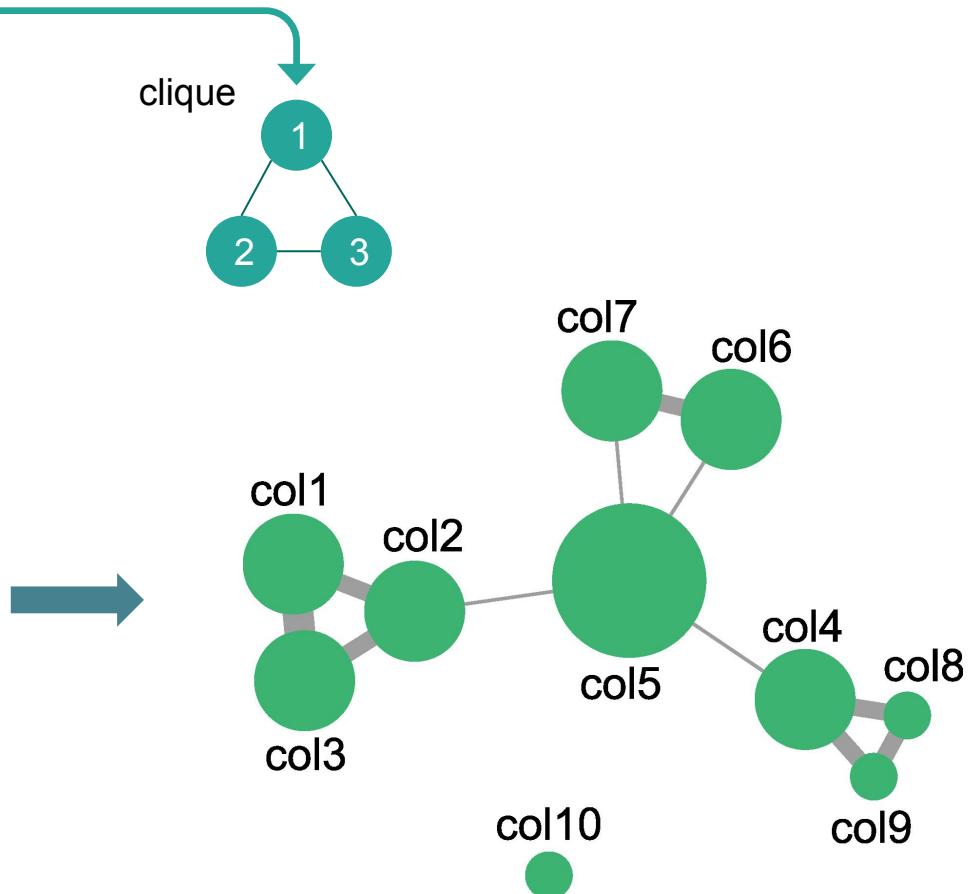
Perspectiva

Co-autoria em registros de espécies é um indicativo do nível de colaboração profissional entre dois coletores



Redes Colaborativas de Coletores (CWN) Construção

id	recordedBy	species
0	col1; col2; col3	sp1
1	col3; col1; col2	sp2
2	col1; col3	sp3
3	col5; col4	sp3
4	col5; col2	sp3
5	col5; col6	sp5
6	col5; col7	sp4
7	col6; col7	sp6
8	col6; col7	sp7
9	col4; col8; col9	sp4
10	col4; col9; col8	sp5
11	col10	sp6
12	col10	sp6

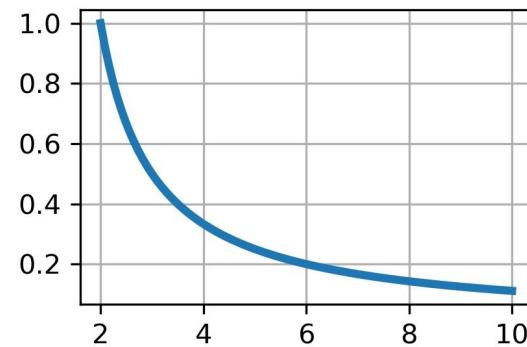


Redes Colaborativas de Coletores (CWN)

Ponderação de arestas

Modelo Hiperbólico

$$w_{(i,j)} = \sum_k \frac{\delta_i^{(k)} \delta_j^{(k)}}{(n^{(k)} - 1)}$$



Para cada registro k , com tamanho de equipe $n^{(k)}$

- $\delta_i^{(k)} = 1$ se o coletor i está no registro k
- $\delta_i^{(k)} = 0$ caso contrário

Atos colaborativos envolvendo muitos coletores tendem a gerar enlaces mais fracos entre os envolvidos [Newman, 2001]

obs: A regra só se aplica para registros com tamanho de equipe $n^{(k)} >= 2$

Estudo de Caso

O Herbário UB Universidade de Brasília



Uma coleção de referência para
a flora do bioma Cerrado

O dataset

- Registros de ocorrência de espécies
- Obtido através do portal GBIF
- Total de 185301 registros
- Apenas 152379 (82.2%) com resolução taxonômica em espécie

O Herbário UB

Caracterização Geral

Taxonômica

- Majoritariamente composto por plantas (Reino *Plantae*)
- Plantas vasculares (Filo *Tracheophyta*) são prevalentes, seguido de musgos (Filo *Bryophyta*) e algas (Filos *Ochrophyta* e *Charophyta*)

	num of taxa	top-4 taxa	num of records	% of records
kingdom	5	Plantae Chromista Fungi Bacteria	179218 4204 1391 342	96.72 2.27 0.75 0.18
phylum	12	Tracheophyta Bryophyta Ochrophyta Charophyta	153589 16485 4133 3500	82.89 8.90 2.23 1.89

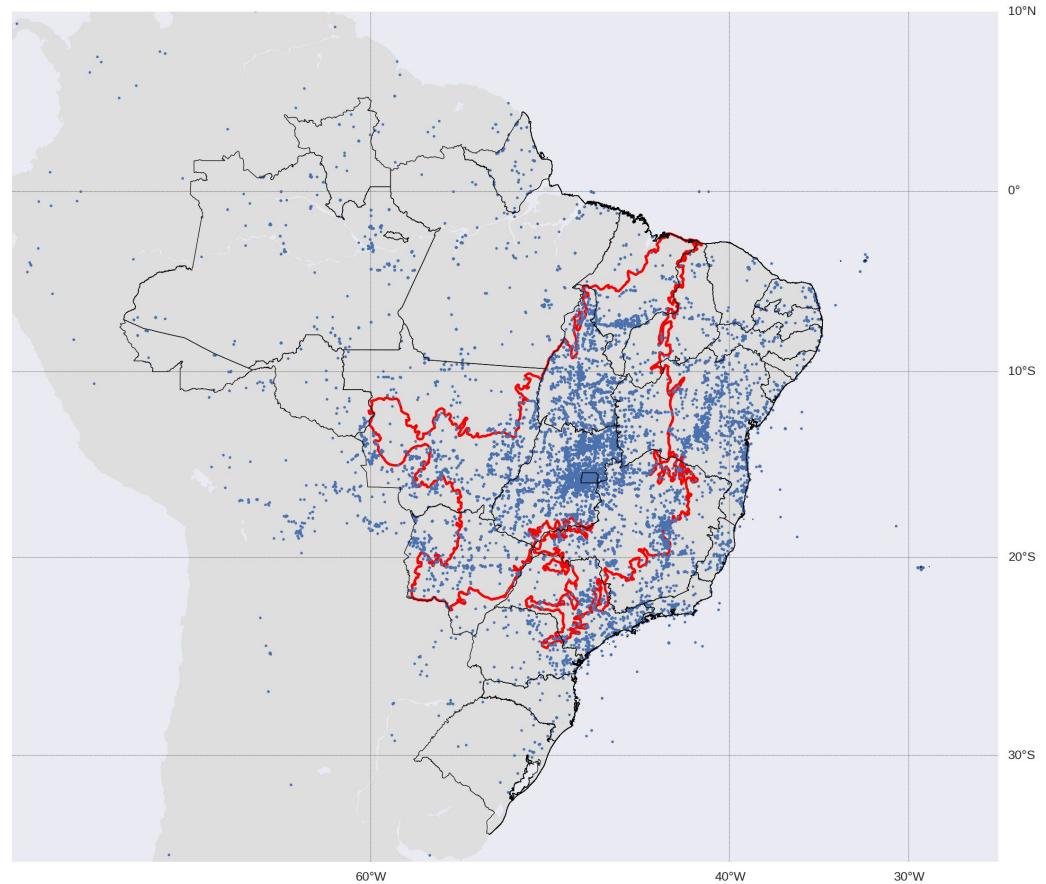
O Herbário UB Caracterização Geral

Geográfica

Total de 94.5% dos registros coletados no Brasil

Dentre os registros no Brasil:

- DF: 30.56%
- GO: 22.10%
- MG: 14.51%
- MT: 7.86%

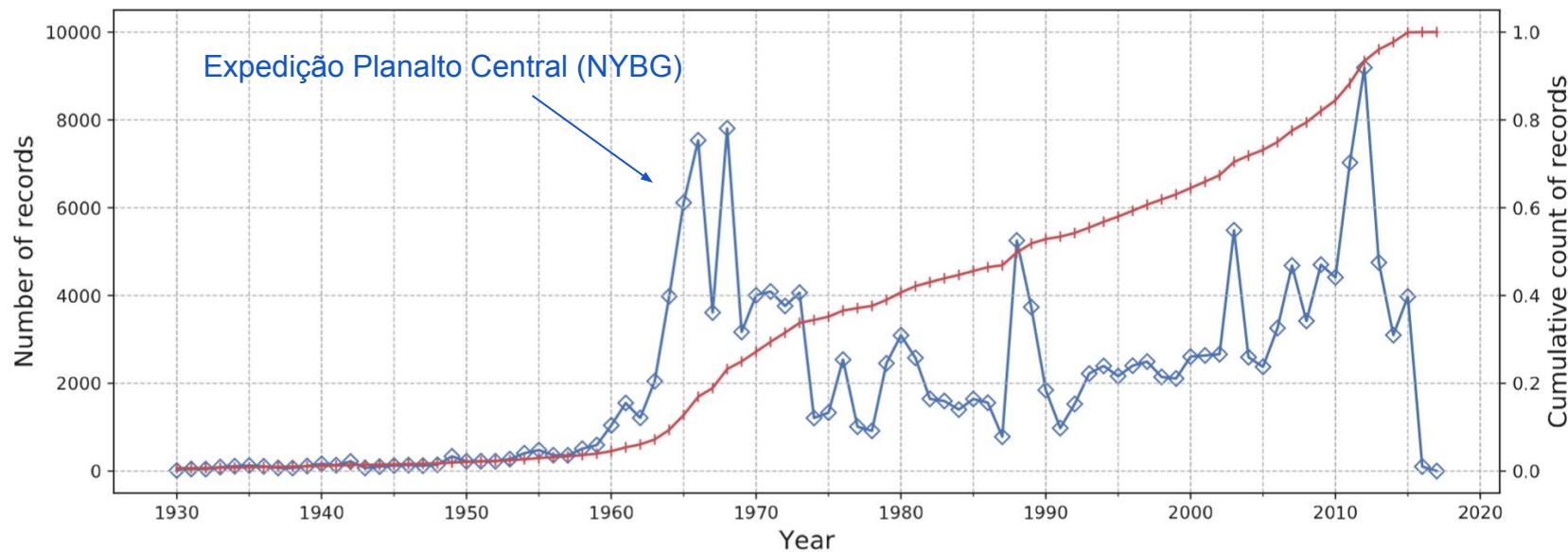


O Herbário UB

Caracterização Geral

Temporal

- Coletas mais intensas a partir de 1960 (96.11% dos registros)
- Dois picos de atividades mais significativas: ~1963-74 , ~2010-15



O Herbário UB Rede Espécie-Coletor (SCN)

Características gerais

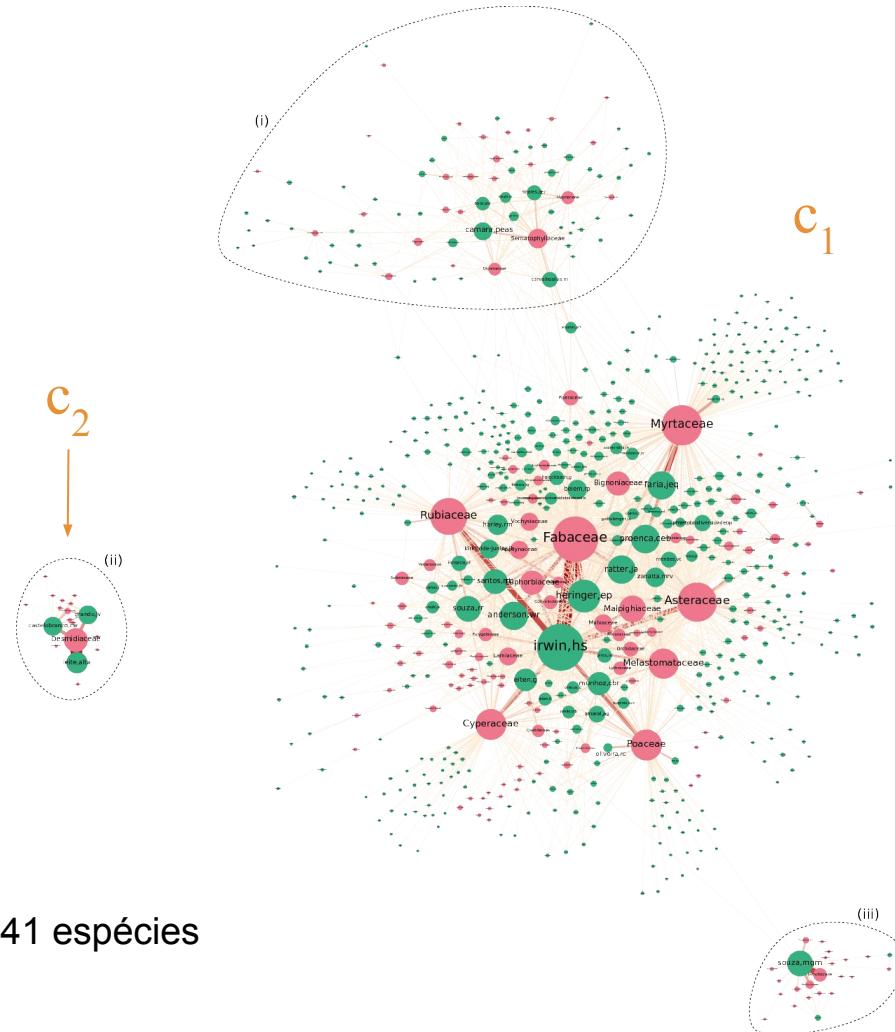
- 6768 coletores e 15344 espécies
- 142647 arestas
- Grau médio:
 - ◆ 21.08 (coletores)
 - ◆ 9.3 (espécies)

Componente gigante (C_1)

- Contém 94% dos coletores
- Contém 95% das espécies

Componentes conectados

- Total de 350 componentes
- O maior deles é C_2 : 3 coletores e 141 espécies



O Herbário UB Rede Espécie-Coletor (SCN)

Distribuição de Grau

Espécies (à esquerda)

Cauda longa, aproximada por uma lei de potência com corte exponencial ($\alpha = 1.38$; $\lambda = 0.0014$)

$$p(k) \sim e^{-\lambda k} k^{-\alpha}$$

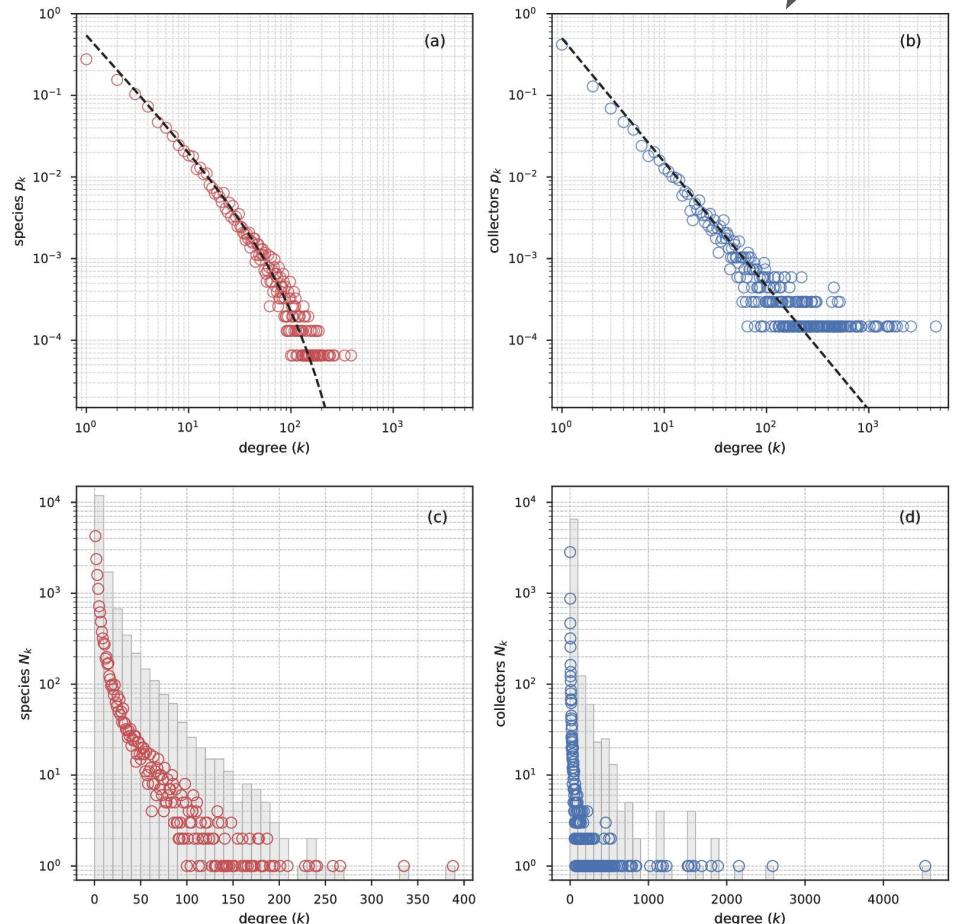
Muitas espécies registradas por poucos coletores, poucas espécies registradas por muitos coletores

Coletores (à direita)

Cauda longa, aproximada por um lei de potência ($\alpha = 1.52$)

$$p(k) \sim k^{-\alpha}$$

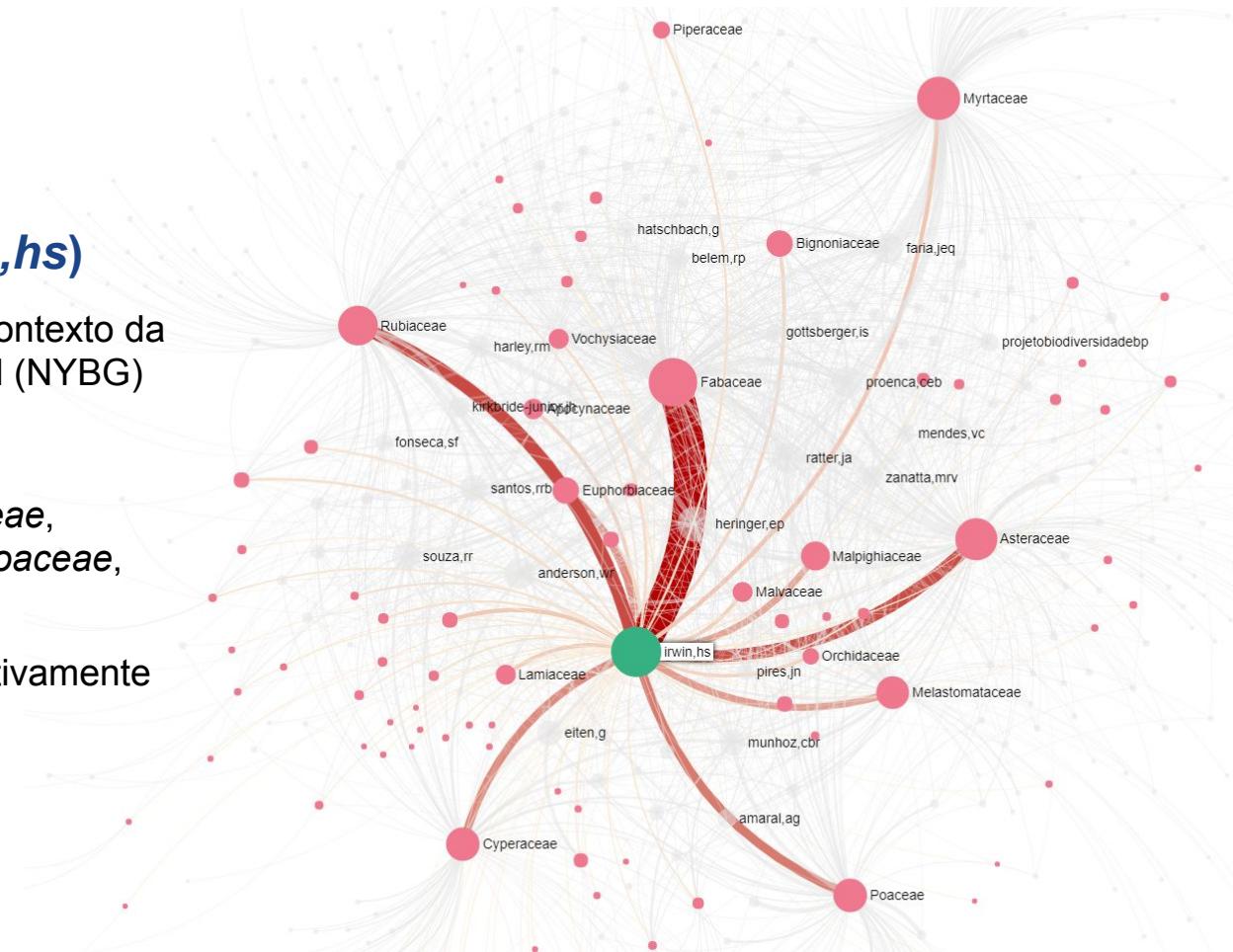
Muitos coletores registram poucas espécies, poucos coletores registram muitas espécies



O Herbário UB Rede Espécie-Coletor (SCN) Coletores

Howard S. Irwin (*irwin,hs*)

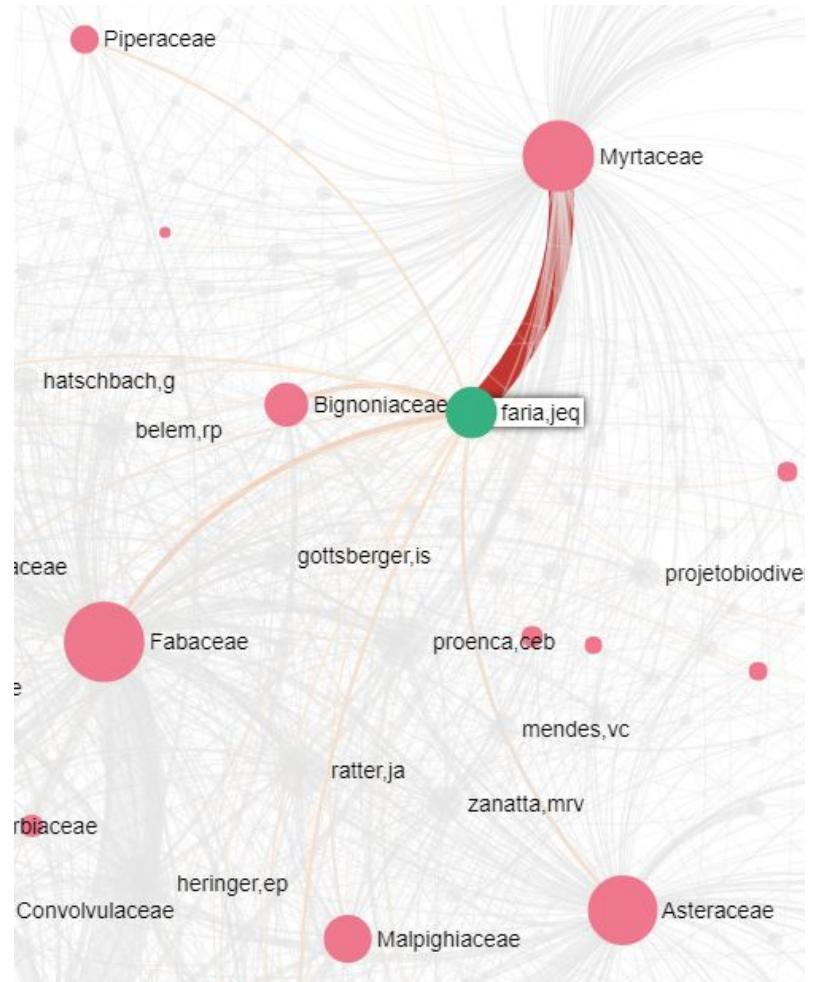
- Contribuiu com o UB no contexto da expedição Planalto Central (NYBG)
- Perfil mais generalista
- Principais famílias: *Fabaceae*, *Rubiaceae*, *Asteraceae*, *Poaceae*, *Cyperaceae*
- Família *Myrtaceae* for relativamente pouco coletada por Irwin



O Herbário UB Rede Espécie-Coletor (SCN) Coletores

Jair E. Faria (*faria,jeq*)

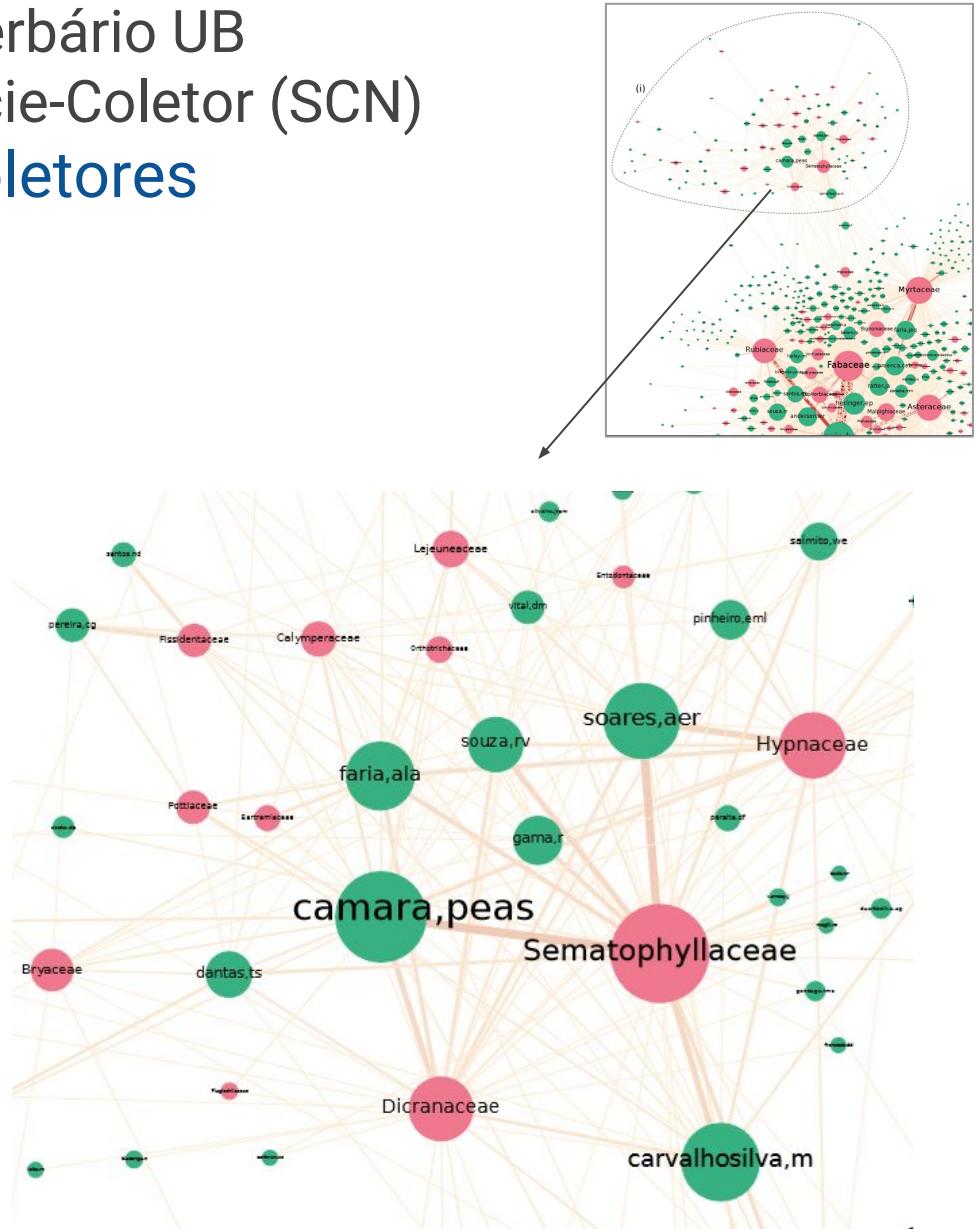
- Mestrado e Doutorado em Botânica na UnB
- Mais especializado na família *Myrtaceae* (31% de suas coletas)



https://lncc-netsci.github.io/pedrocs/networks/ub_scn/

O Herbário UB Rede Espécie-Coletor (SCN) **Coletores**

Laboratório de Criptógamas UnB



Paulo E. Câmara (camara,peas)

- Pesquisador e Professor Adjunto na UnB, desde 2000

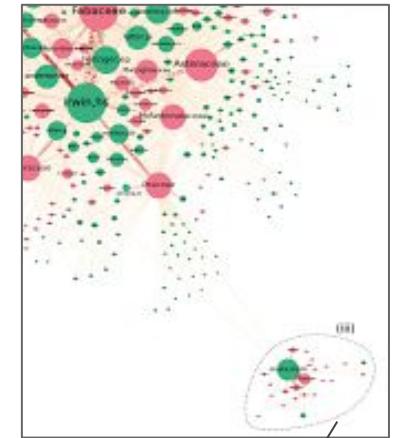
Micheline C. Silva (carvalhosilva,m)

- Pesquisadora e professora colaboradora na UnB; adjunta na UFVJM
 - Interesse em briófitas e também angiospermas (*Piperaceae*)

Abel E. R. Soares (soares, aer)

- Mestrado e Doutorado em Botânica na UnB

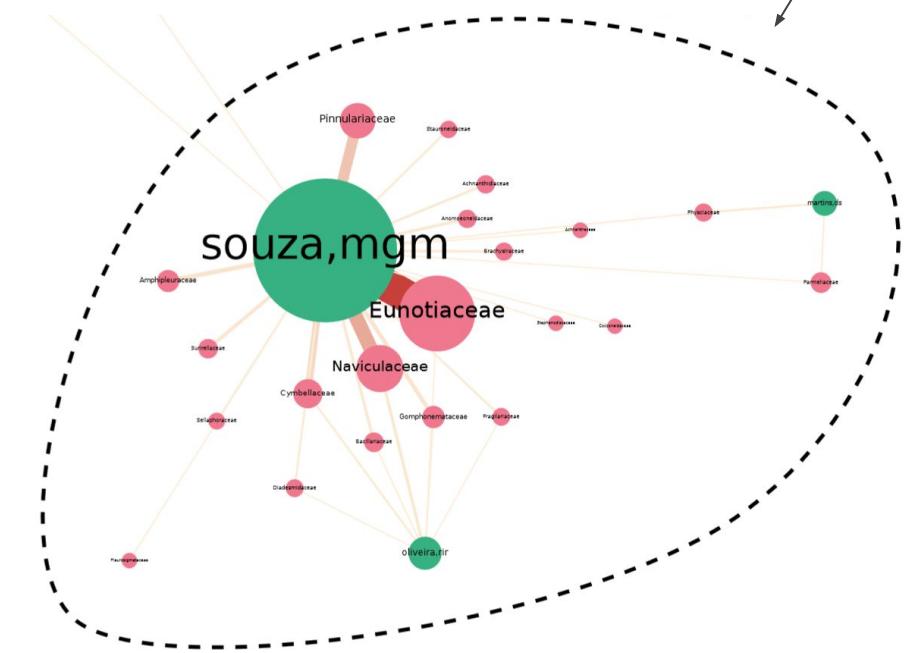
O Herbário UB Rede Espécie-Coletor (SCN) Coletores



Laboratório de Criptógamas UnB

Maria das Graças M. Souza (*souza,mgm*)

- Pesquisadora e Professora na UnB, desde 1993
- Especialista em algas continentais, principalmente diatomáceas (ex. famílias *Eunotiaceae*, *Naviculaceae*, *Pinnulariaceae*, ...)



O Herbário UB Rede Espécie-Coletor (SCN) Coletores

Coletores de algas continentais

Christina W. Castelo Branco (*castelobranco,cw*)

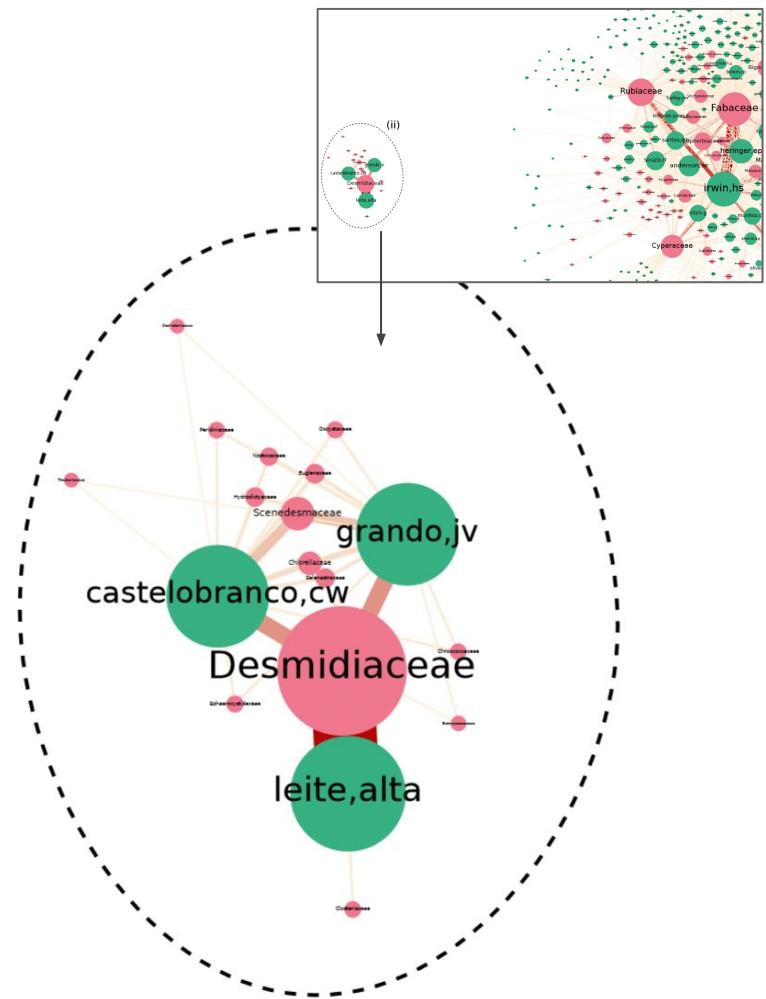
- Mestrado em Ecologia na UnB: comunidade planctônica e qualidade de água no Lago Paranoá
- Algas verdes (principalmente *Desmidiaceae* e *Scenedesmaceae*)

João V. Grando (*grando,jv*)

- Mestrado em Ecologia na UnB: aspectos alimentares da comunidade de peixes no Lago Paranoá
- Interesse igual '*castelobranco,cw*'

Ana Lúcia T. A. Leite (*leite,alta*)

- Mestrado em Ecologia na UnB: família *Desmidiaceae*

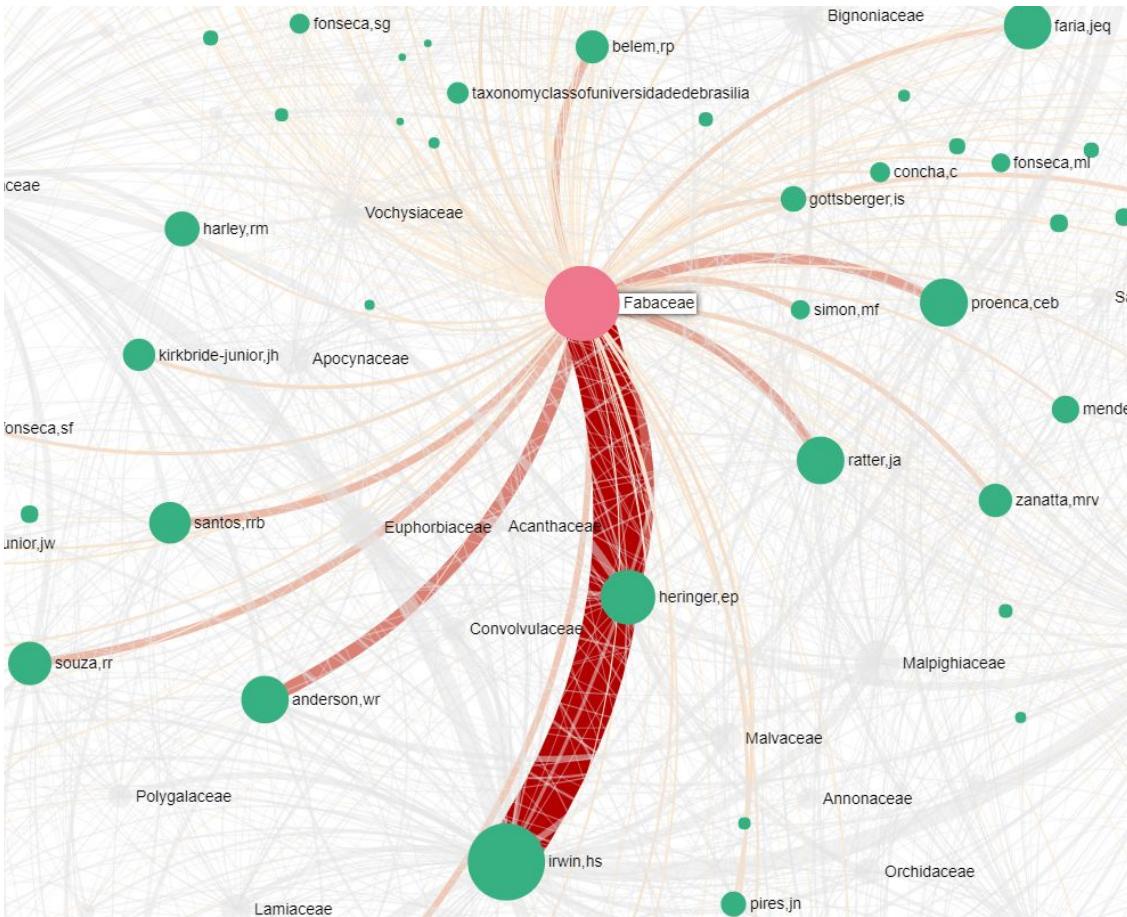


O Herbário UB Rede Espécie-Coletor (SCN) Famílias



Fabaceae

- Uma das maiores famílias de plantas no Cerrado
- A família mais central no modelo: amplamente amostrada por muitos coletores, dentre os quais:
 - Howard Irwin
 - Ezechias Paulo Heringer
 - William R. Anderson
 - Carolyn E. B. Proença
 - James A. Ratter
 - Raimundo Souza
 - Raimundo dos Santos
 - Romeu P. Belém

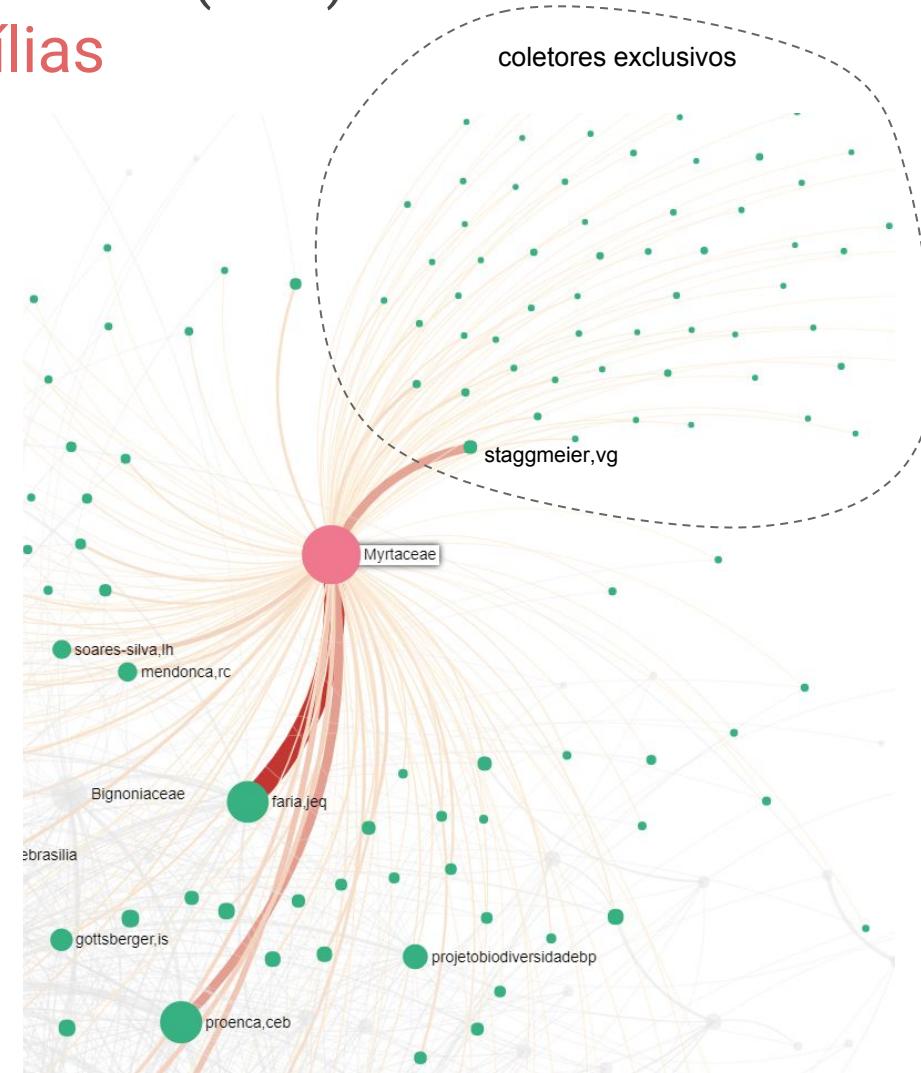




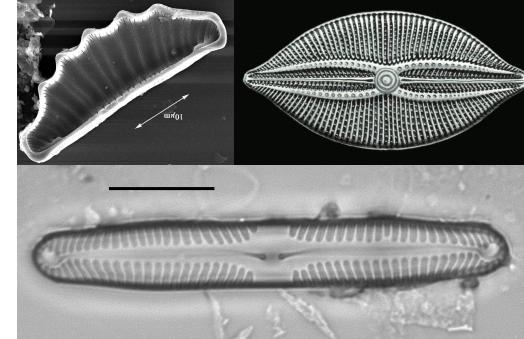
O Herbário UB Rede Espécie-Coletor (SCN) Famílias

Myrtaceae

- A segunda família mais representativa no modelo
- Coletada por um grupo mais **distinto** de coletores (embora também por **generalistas**):
 - Jair E. Faria
 - Vanessa G. Staggmeier
 - Carolyn E. B. Proença
 - Lúcia Helena Silva
- Relativamente pouco coletada por Howard Irwin



O Herbário UB Rede Espécie-Coletor (SCN) Famílias



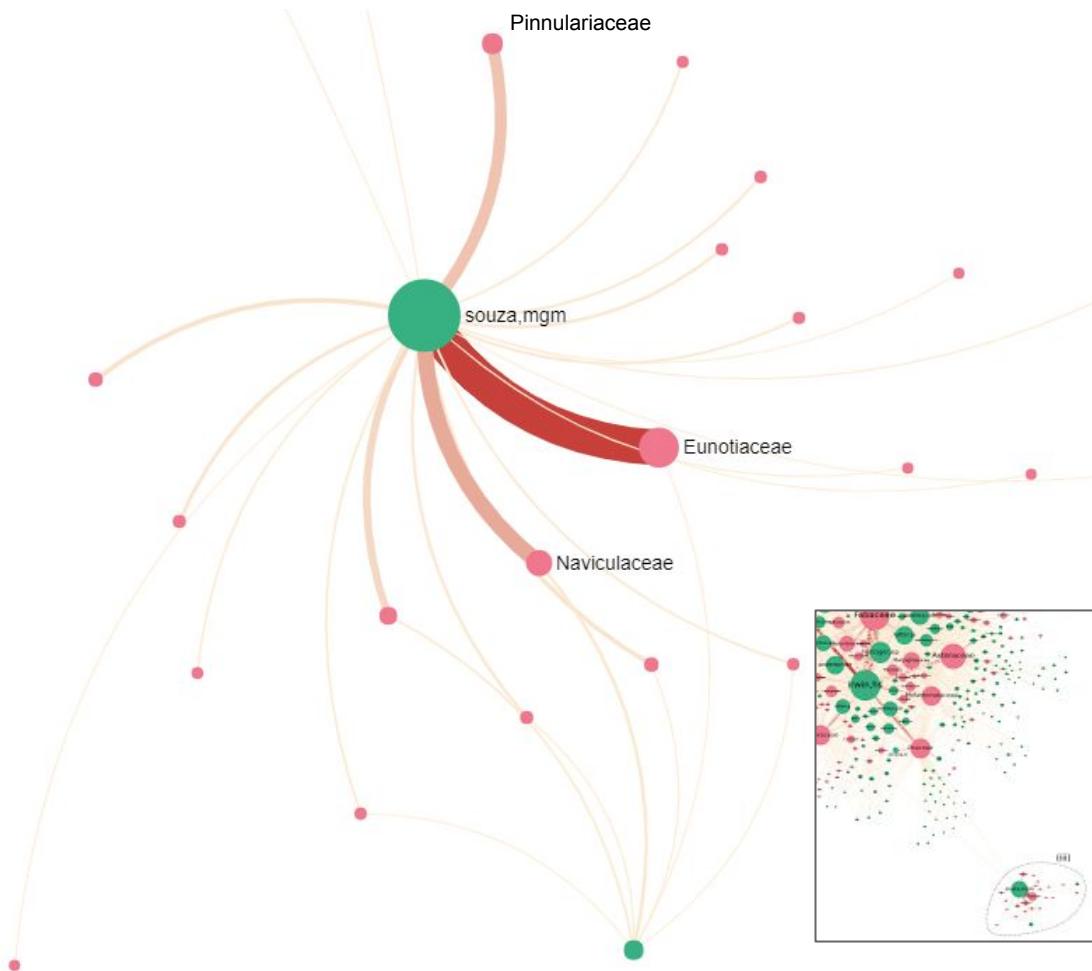
Diatomáceas

Família **Eunotiaceae**

Família **Naviculaceae**

Família **Pinnulariaceae**

- Organismos microscópicos
- Praticamente coletadas apenas por Maria das Graças M. Souza

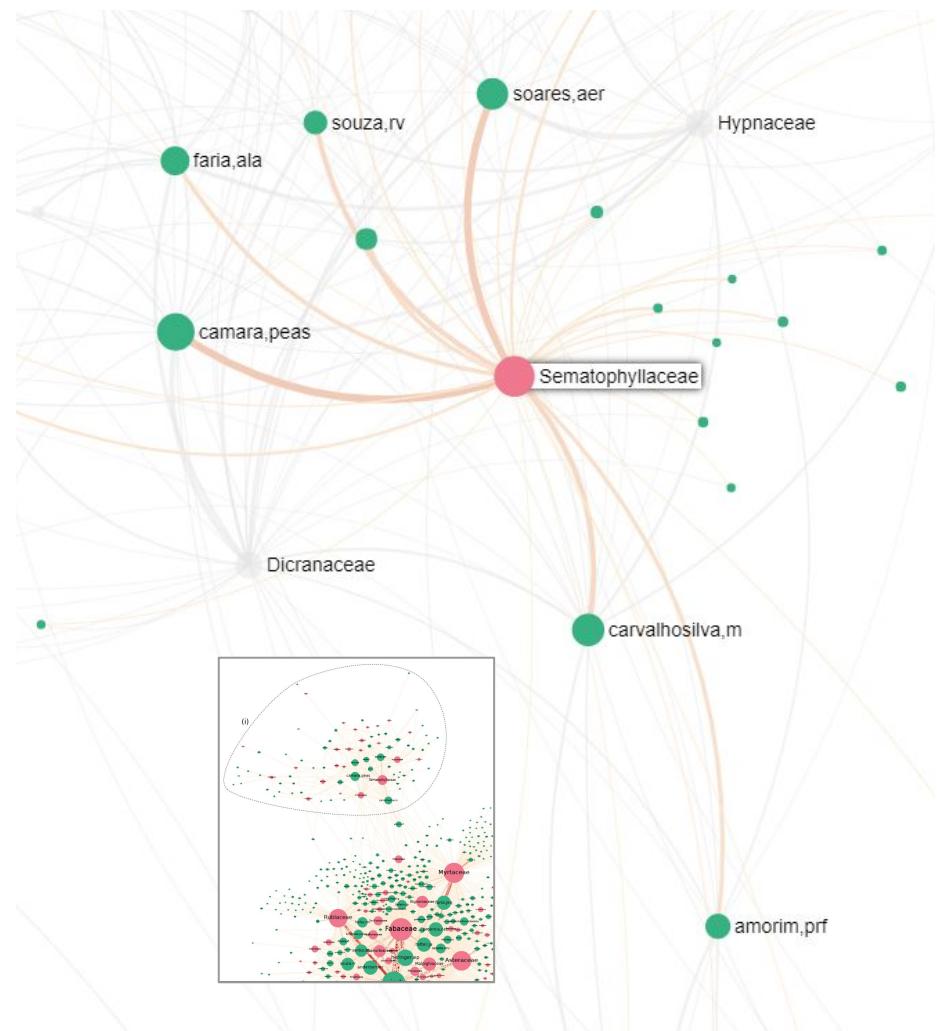


O Herbário UB Rede Espécie-Coletor (SCN) Famílias

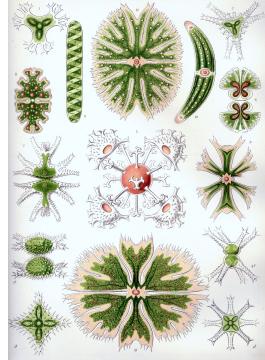


Sematophyllaceae

- Família de musgos
- Praticamente coletados por especialistas, muitos dos quais no Laboratório de Criptógamas:
 - Paulo E. Câmara
 - Abel E. Soares
 - Micheline Carvalho Silva
 - Ronaldo V. Souza
 - Allan L. Faria

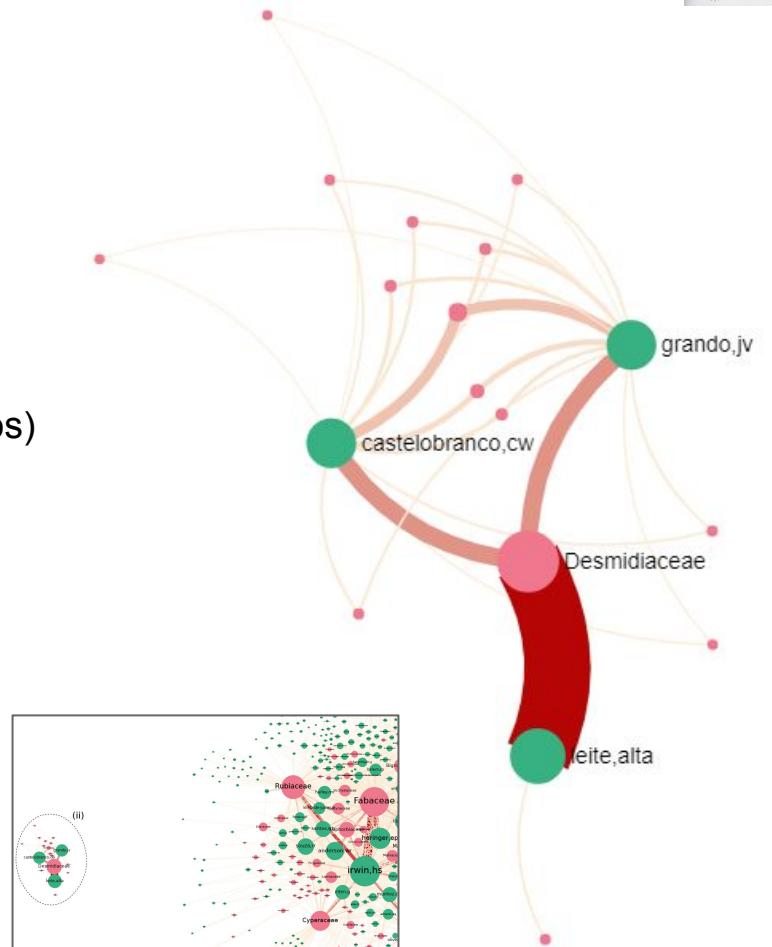


O Herbário UB Rede Espécie-Coletor (SCN) Famílias



Desmidiaceae

- Família de algas verdes (organismos microscópicos)
- Coletadas por apenas três coletores, todos extremamente especializados :
 - Ana Lúcia T. Leite
 - João V. Grando
 - Chrystina W. Castelo Branco



O Herbário UB Rede Espécie-Coletor (SCN) Projeção: coletores

Arestas representam **similaridade** entre coletores, em termos de suas *species bags* (sim ≥ 0.8)

Coletores isolados são distintos em termos de seus interesses taxonômicos (ex ‘souza,mgm’, ‘leite,alta’)

Comunidades de interesse.

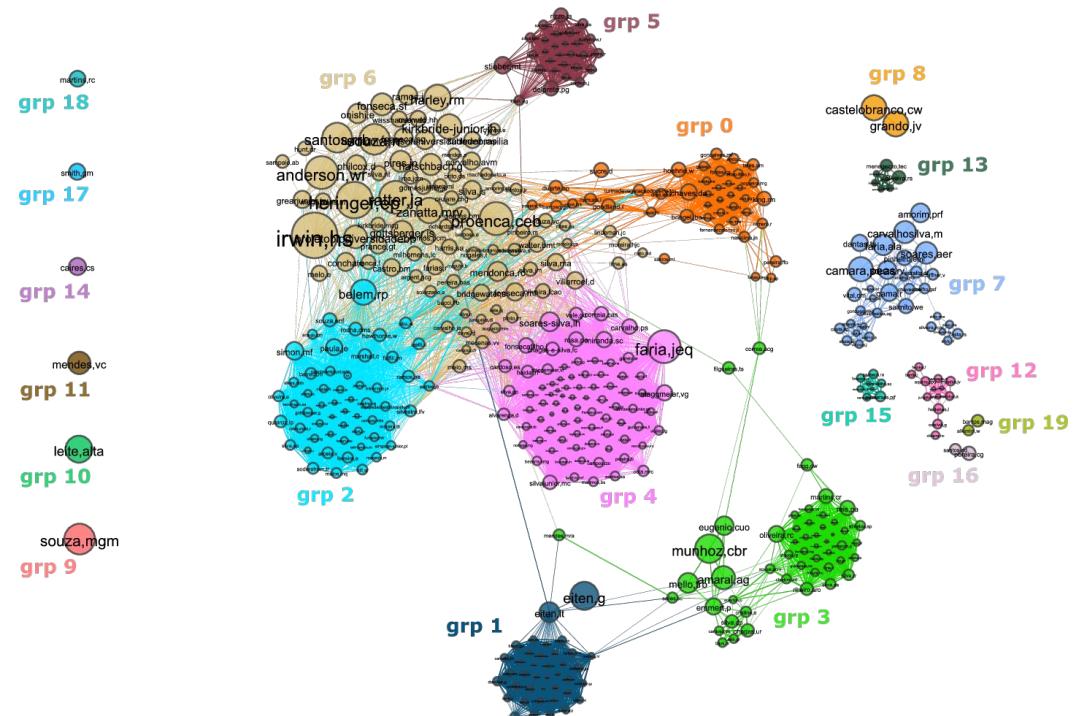
Coletores com interesses taxonômicos similares

grp 6
Coletores generalistas

grp 7 e grp 12
Coletores de briófitas

grp 3
Coletores de plantas herbáceas

grp 0
Coletores especialistas em Asteraceae



O Herbário UB Rede Espécie-Coletor (SCN)

Projeção: espécies

Arestas representam **similaridade** entre espécies, em termos de seus *quorum* ($\text{sim} \geq 0.8$)

Espécies isoladas são coletadas por coletores bastante específicos (ex. famílias *Myrtaceae* e *Poaceae*)

Comunidades contêm espécies que são coletadas por grupos similares de coletores. Este é um agrupamento que distingue bem os interesses dos coletores.

grp 0

Famílias coletadas por uma ampla gama de coletores (maioria generalistas)

grp 4

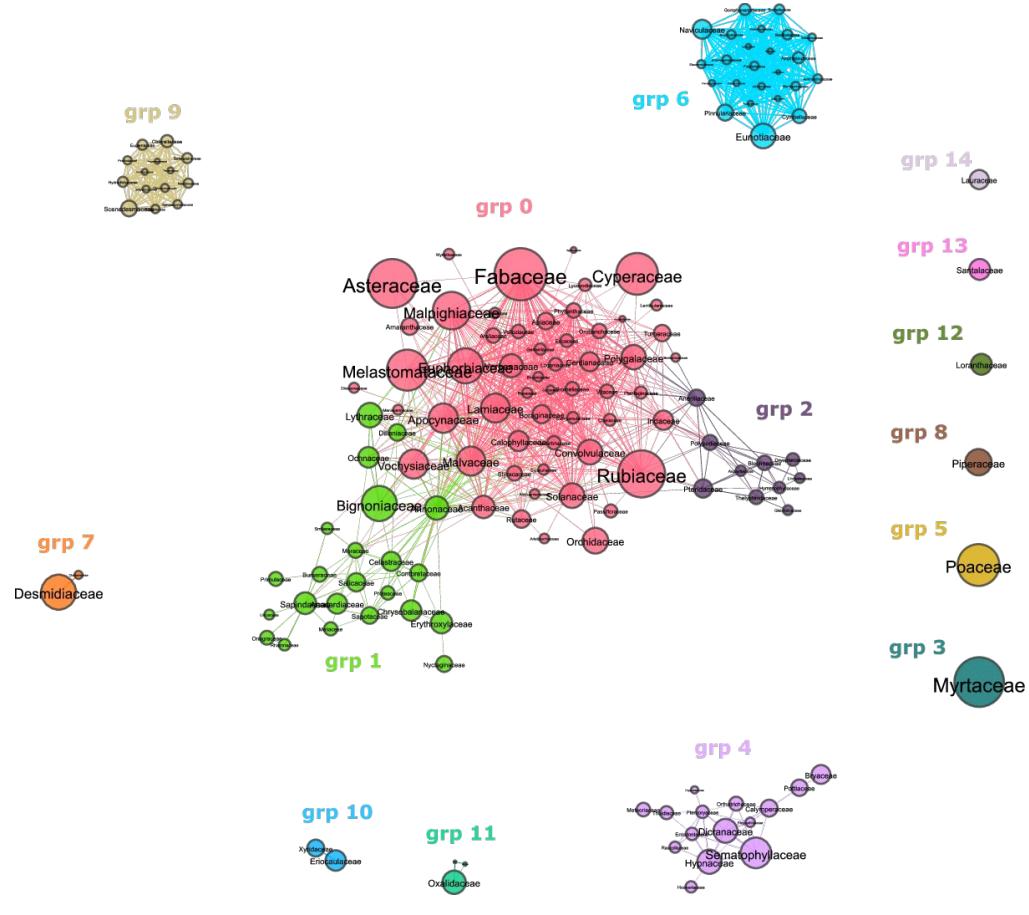
Famílias de briófitas

grp 2

Famílias de samambaias (pteridófitas)

grp 7 e grp 9

Famílias de algas verdes



O Herbário UB Rede Colaborativa de Coletores (CWN)

Características gerais

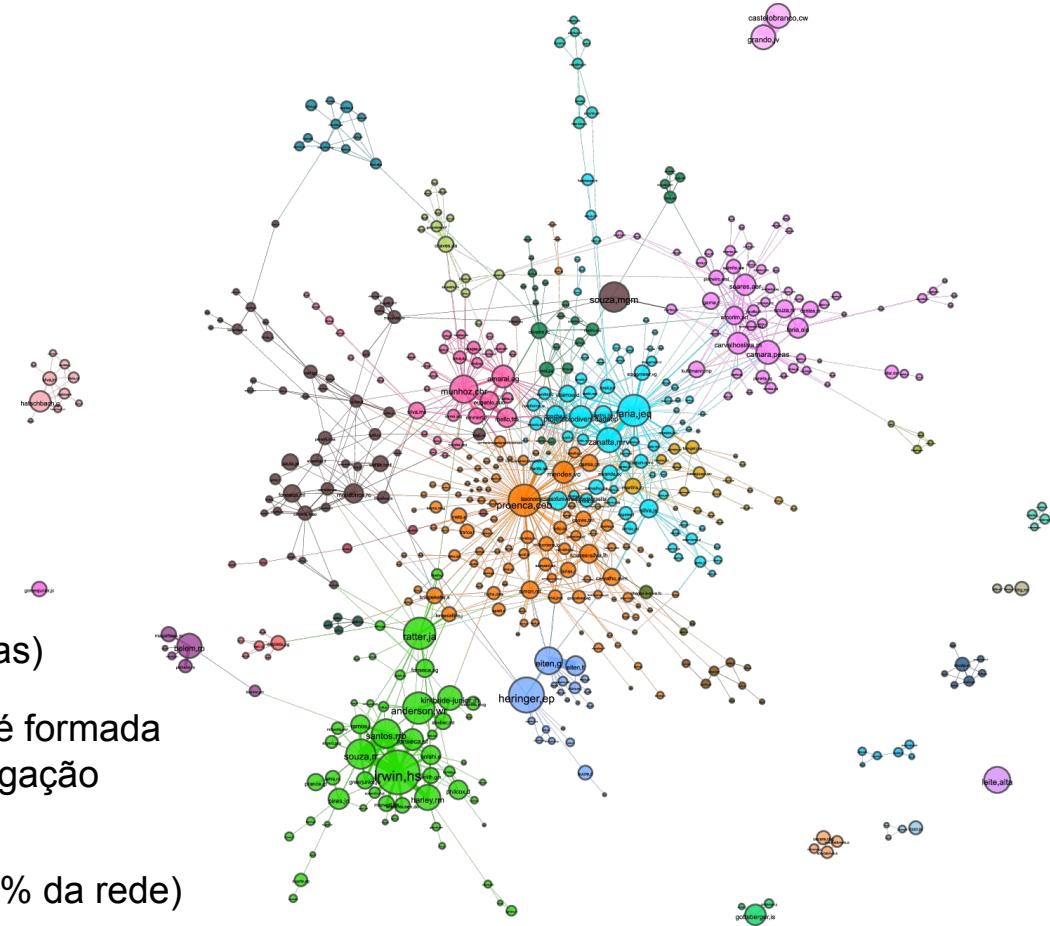
- 6768 coletores
 - 10391 arestas
 - Grau médio $\langle k \rangle = 3.07$

Componente gigante (c_1)

- Contém 46% dos coletores

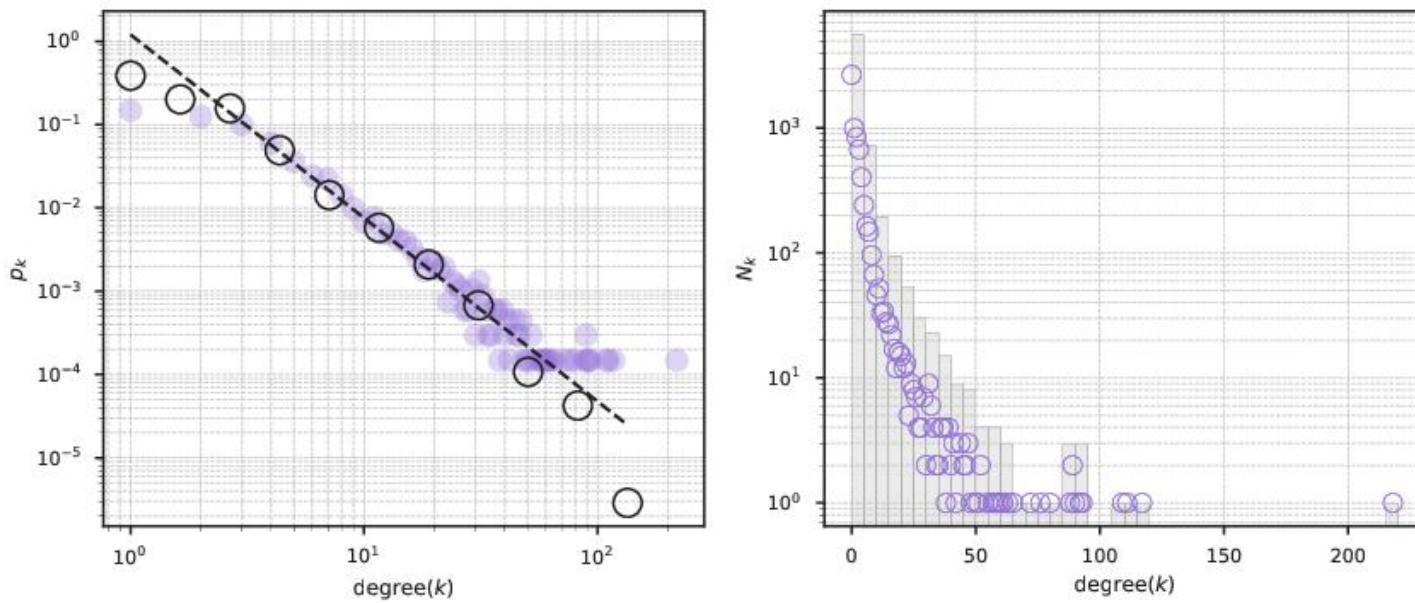
Componentes conectados

- Total de 2990 componentes (ou ilhas)
 - Apenas 318 das ilhas (11% delas) é formada por coletores com ao menos uma ligação colaborativa
 - 2672 coletores *individualistas* (39.5% da rede)



O Herbário UB Rede Colaborativa de Coletores (CWN)

Distribuição de Grau



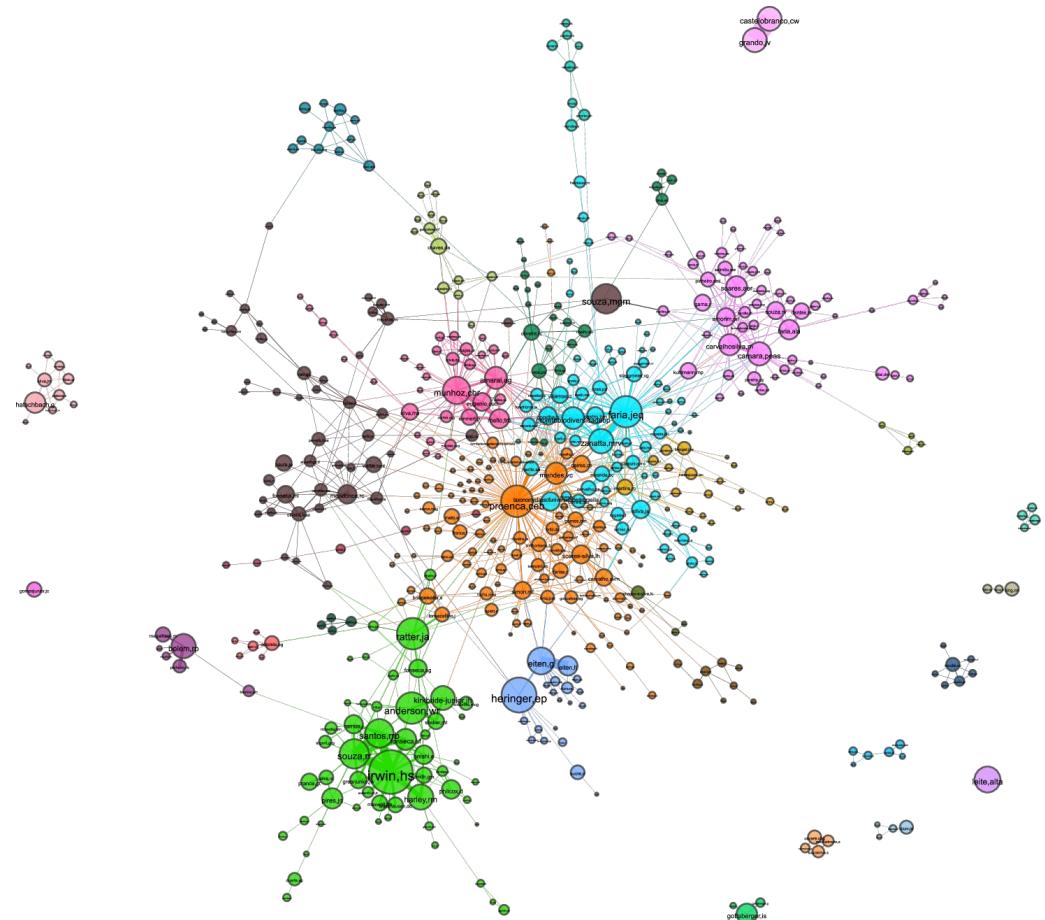
Cauda longa, aproximada por uma lei de potência ($\alpha = 2.2$) $p(k) \sim k^{-\alpha}$

Muitos coletores pouco colaborativos coexistem com poucos coletores muito colaborativos

O Herbário UB Rede Colaborativa de Coletores (CWN) Comunidades de colaboração

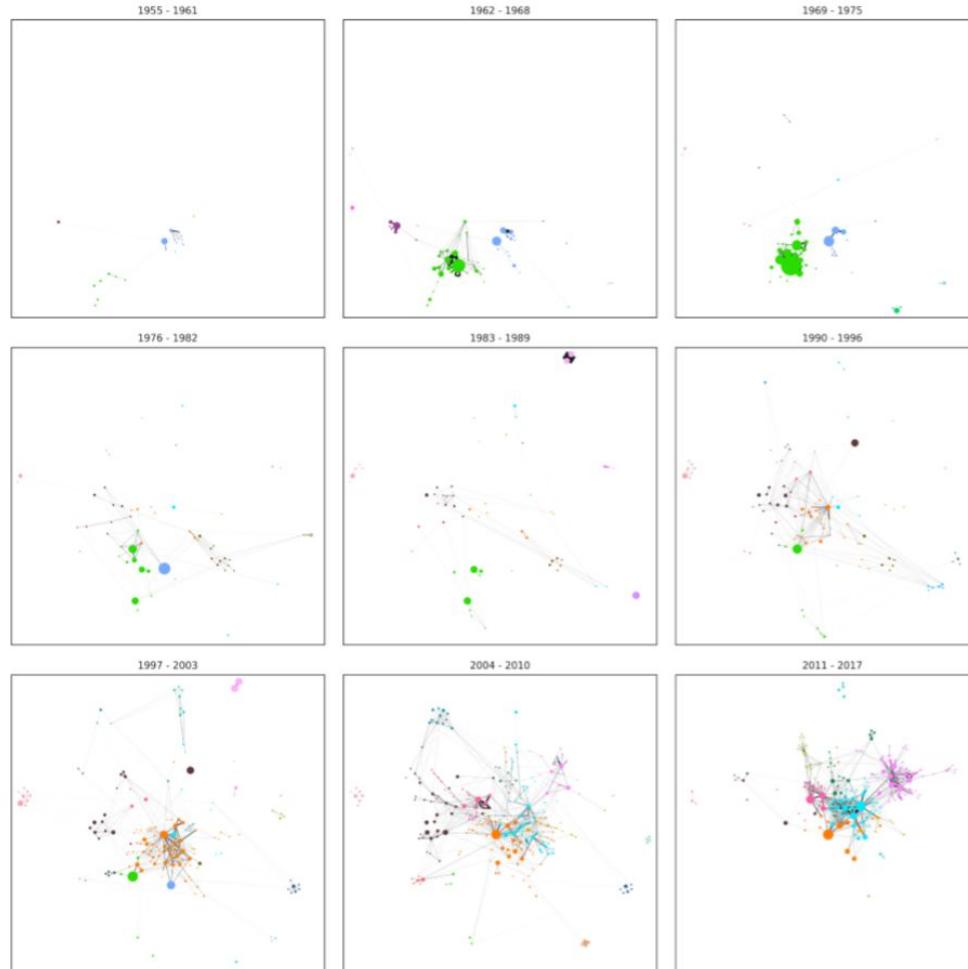
Agrupam coletores que **trabalham juntos** em campo

- Equipes com interesses similares;
 - Equipes com interesses distintos;
 - Equipes separadas temporalmente.



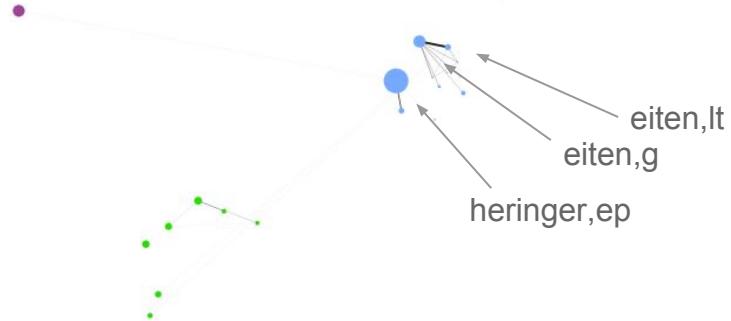
O Herbário UB Rede Colaborativa de Coletores (CWN)

Evolução Temporal



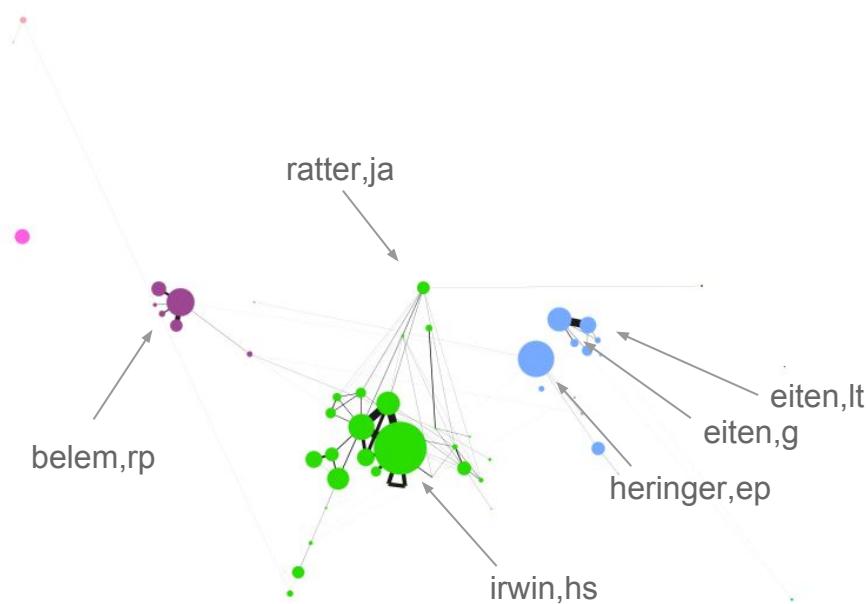
Primeira Geração

1955 - 1961



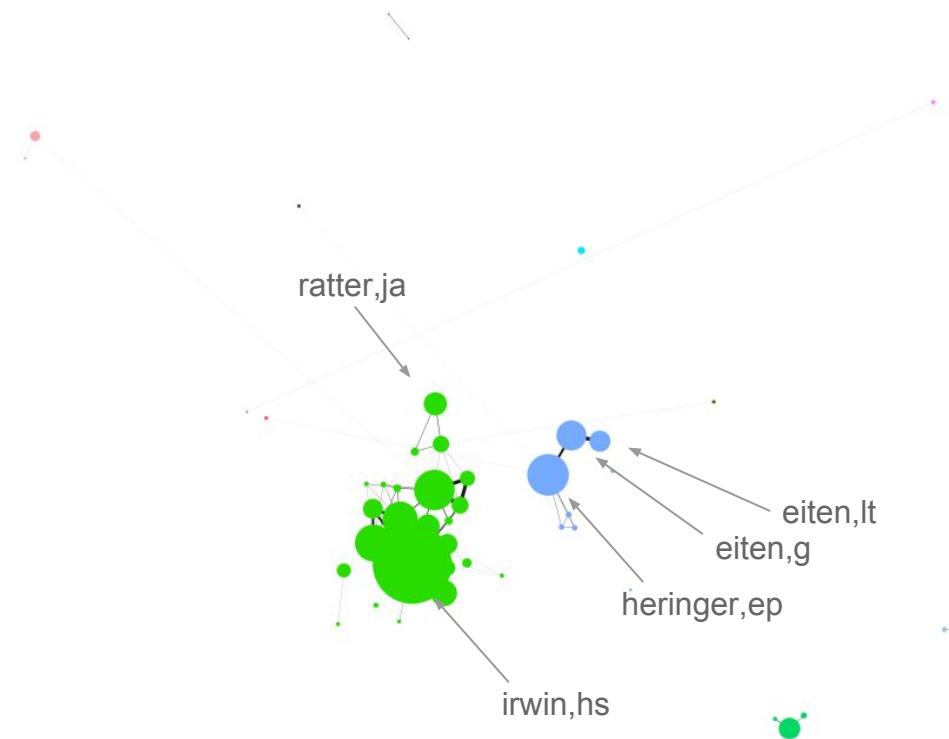
Primeira Geração

1962 - 1968



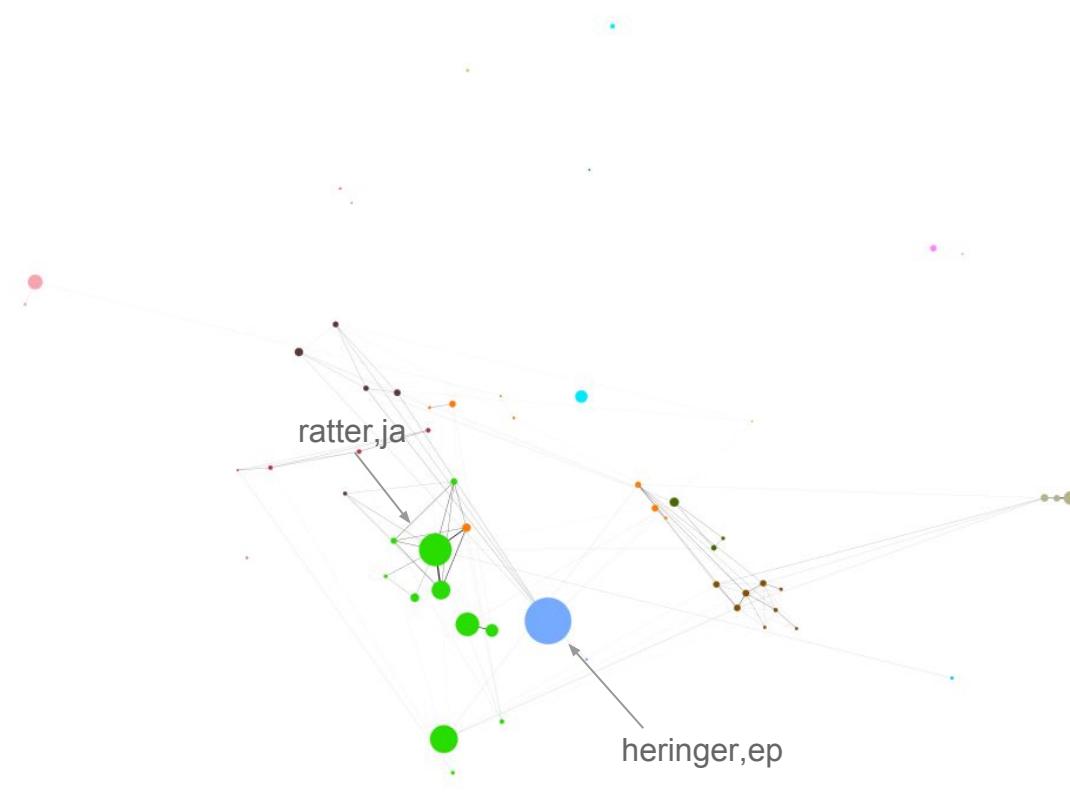
Primeira Geração

1969 - 1975



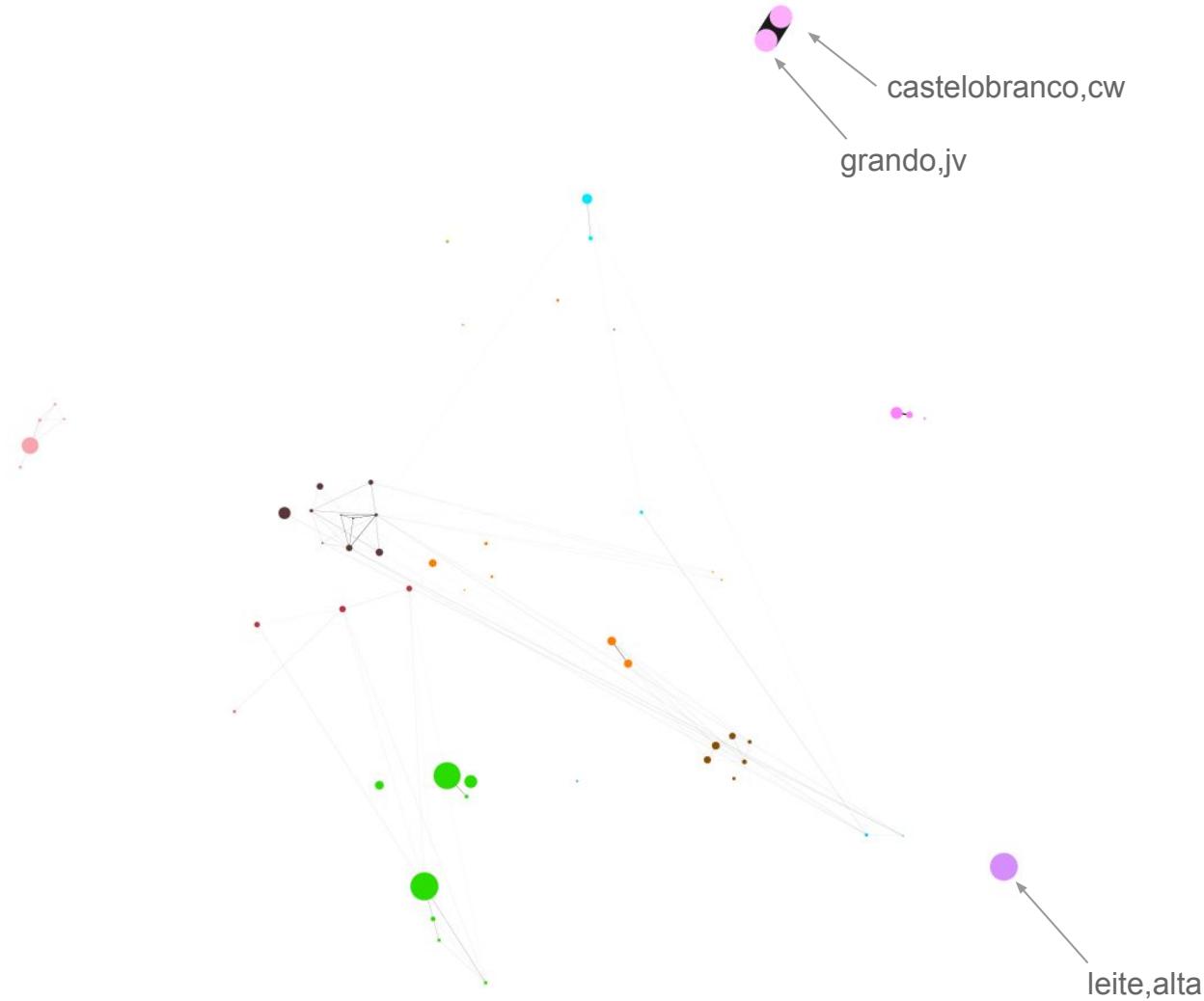
Primeira Geração

1976 - 1982



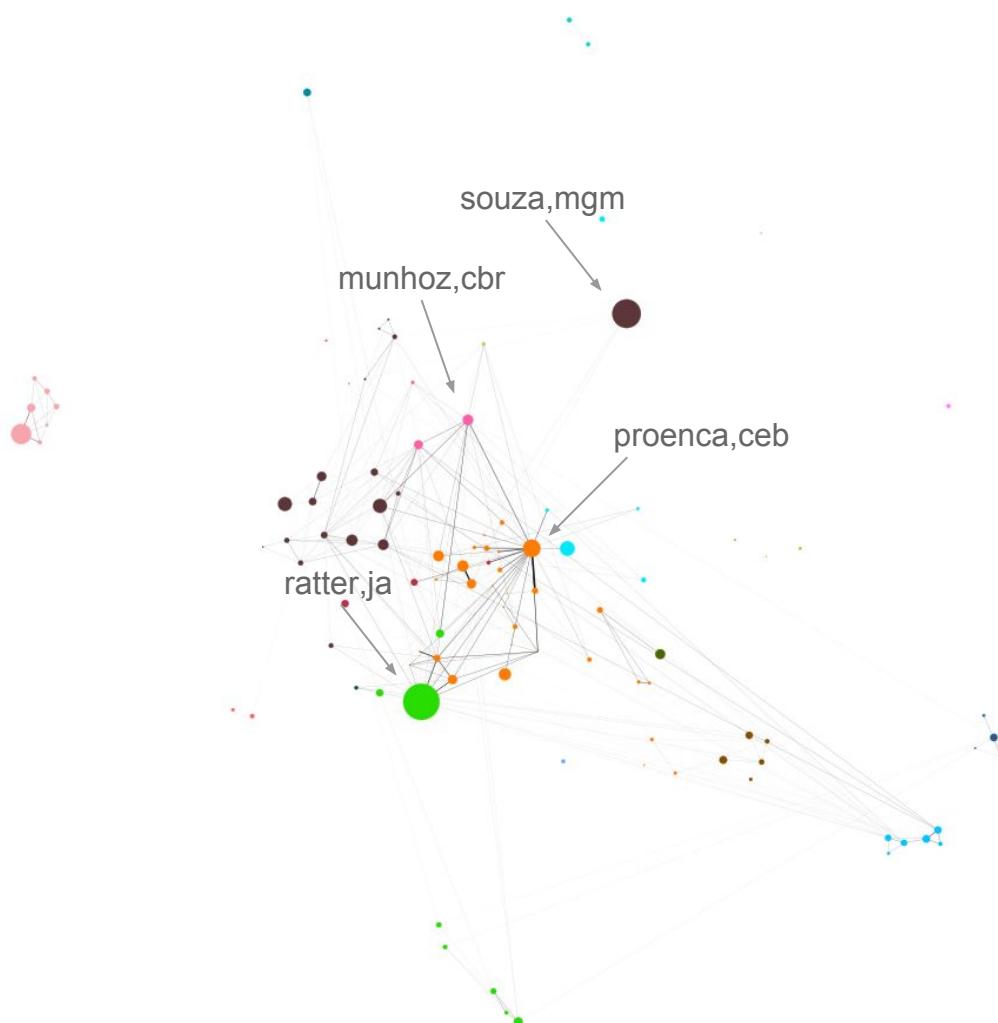
Segunda Geração

1983 - 1989



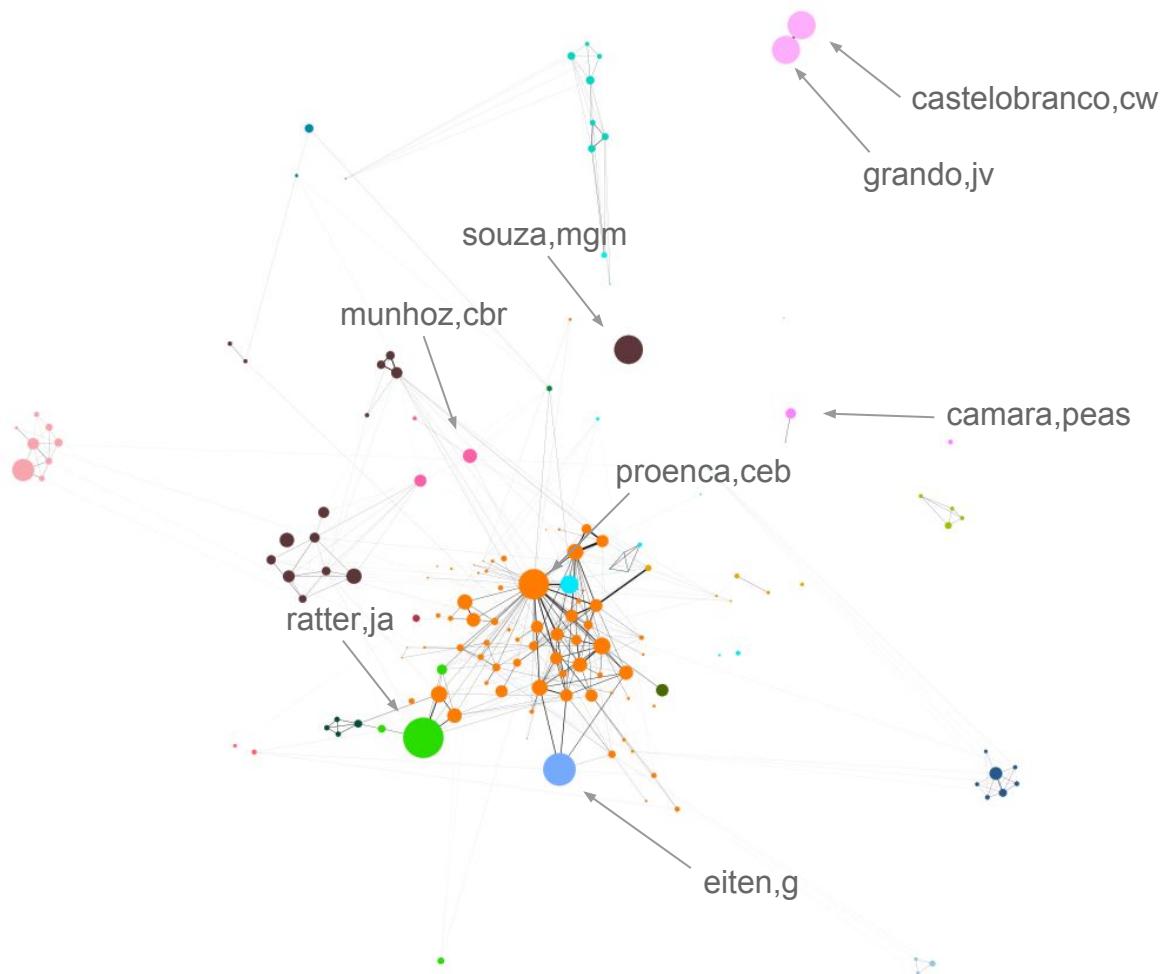
Segunda Geração

1990 - 1996



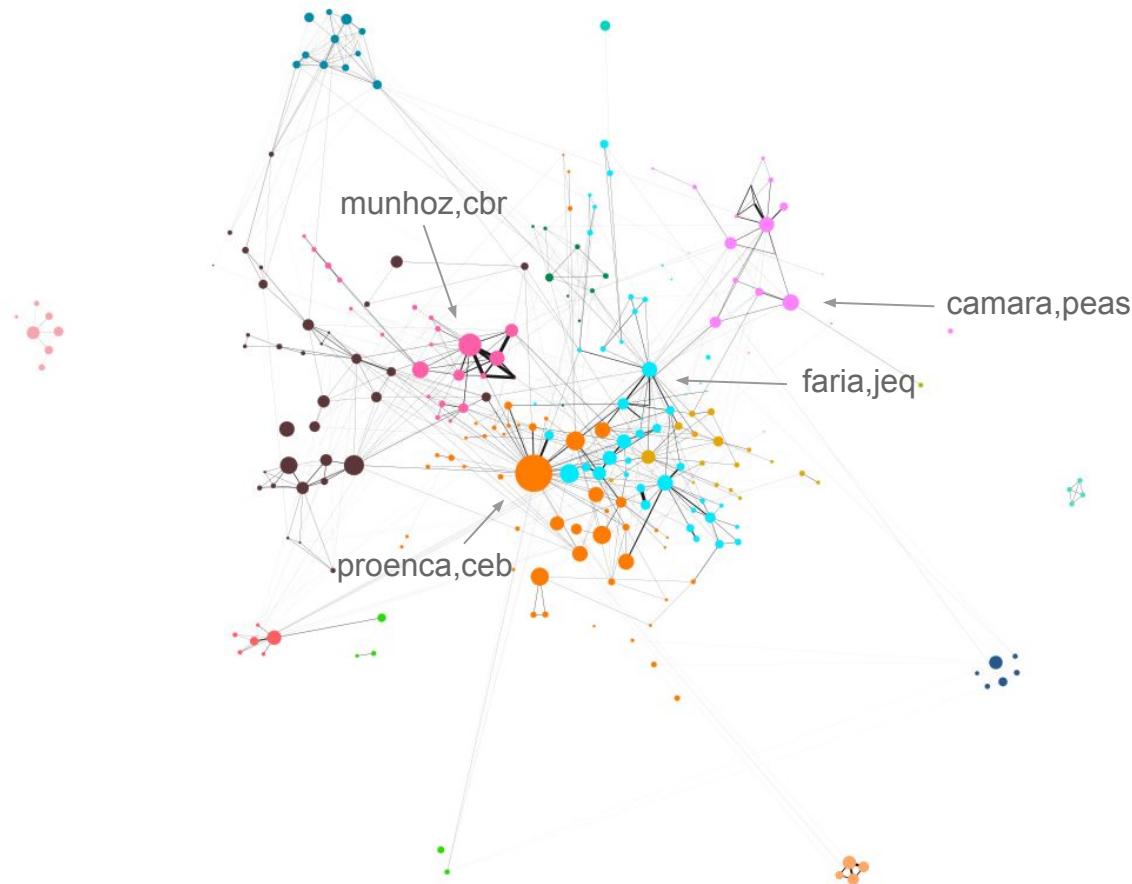
Segunda Geração

1997 - 2003



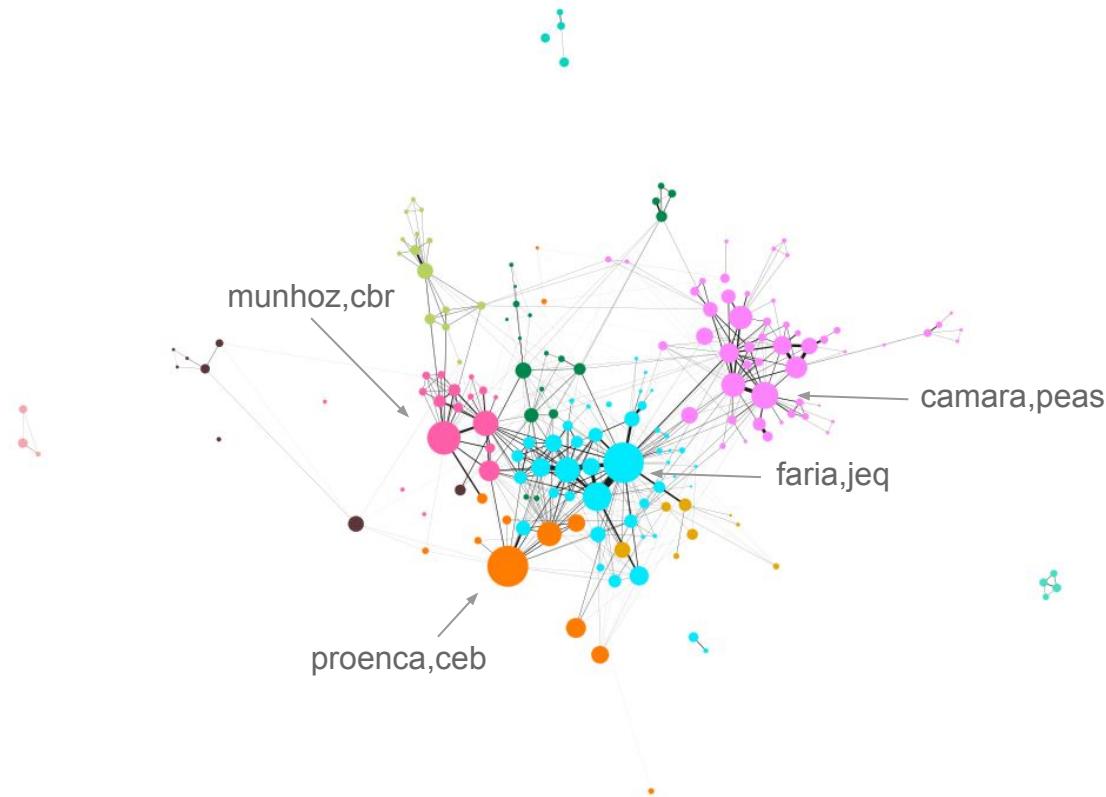
Terceira Geração

2004 - 2010



Terceira Geração

2011 - 2017



Conclusão Pontos Centrais



- Montagem das CBs resulta de um complexo arranjo de interesses e comportamentos individuais dos coletores
- Composição de CBs tende a refletir os interesses dos coletores mais produtivos
- Coletores **interagem** e **influenciam** uns aos outros
- Modelos em redes fornecem a base estrutural para representar e investigar este sistema complexo:
 - **SCNs**: Modelam interesses de **coletores** sobre **espécies**
 - **CWNs**: Modelam relações de colaboração entre **coletores**

Conclusão Limitações dos Modelos



- Qualidade dos modelos depende da qualidade dos dados de ocorrência de espécies usados para construí-los;
- Modelos incorporam a perspectiva de uma única CB;
- Modelos são **estáticos** e **não-espacializados**:
 - Restrições temporais:
 - Colaborações só poderiam ocorrer com sobreposição temporal
 - Links se formariam e quebrariam continuamente
 - Restrições geográficas:
 - Colaborações só poderiam ocorrer com sobreposição espacial
 - Espécies só poderiam ser coletadas se sua área de ocorrência tivesse sido amostrada

Conclusão Perspectivas



1. Perfilamento de coletores

Coletores podem ser classificados em **perfis** distintos:

- Professor, Aluno, Assistente de campo;
- Experiente ou Novato;
- Especialista ou Generalista

Problema de Classificação de Vértice

Iniciando com um subconjunto de vértices (coletores) já rotulados, outros vértices são rotulados de acordo com seus atributos, posição na rede e padrões de associação.

Conclusão Perspectivas



2. Enriquecimento contextual de ocorrências

Perfis de coletores nas equipes poderiam ser utilizados para inferir aspectos contextuais dos registros. Hipoteticamente:

- Professor + vários Alunos -> Disciplina de campo;
- Coletores estrangeiros (fora do C.G. na CWN) -> permuta entre CBs;
- Coletores generalistas -> amostragem generalista
- Coletores especialistas -> amostragem especialista
- Vários coletores experientes com distintas especialidades -> *big survey*;

Conclusão Perspectivas



3. Validação colaborativa de IDs de coletores

Resolver a identidade de coletores nos modelos é um grande desafio

Abordagem: Requisitar que coletores (cuja identidade já foi resolvida):

1. Colaborem com a identificação de entidades na rede com as quais já trabalhou;
2. Propaguem a mensagem para os colegas

Qual o melhor conjunto de vértices para iniciar a mensagem?

Análogo à dinâmica de **contágio** em redes

Conclusão Perspectivas



4. Formação de equipes de especialistas

Como formar uma equipe de coletores para uma determinada tarefa maximizando a eficiência de sua colaboração?

Dois passos:

1. Caracterizar a expertise de coletores (a partir de SCNs)
2. Formar equipes que preencham os requisitos e que tenham previamente colaborado (CWNs)

Aplicações:

- Seleção de colaboradores para a construção de listas vermelhas de espécies
- Seleção de equipes de campo

Conclusão Perspectivas



5. Acurácia nas determinações taxonômicas

O quanto confiável é uma determinação taxonômica dada por um taxonomista?

Modelo análogo à SCN: **Rede Espécie-Determinador**

Modela o interesse (ou experiência) de um taxonomista em relação às espécies que identifica

Score de confiança é associado a cada determinação, com base no perfil do autor

Obrigado!

That's all Folks!

Bibliografia

Daru, B.H. et al. Widespread sampling biases in herbaria revealed from large-scale digitization: Sampling bias in herbarium specimens. 2017. *New Phytologist*, v. 217, n.2, p.939-955.

Lambiotte, R.; Ausloos M. Uncovering collective listening habits and music genres in bipartite networks. 2005. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, v. 72, n. 6.

Newman, M.E.J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. 2001. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, v.64, n. 1.

Redes Espécie-Coletor (SCN) Projeções

Projeção sobre S_{sp}

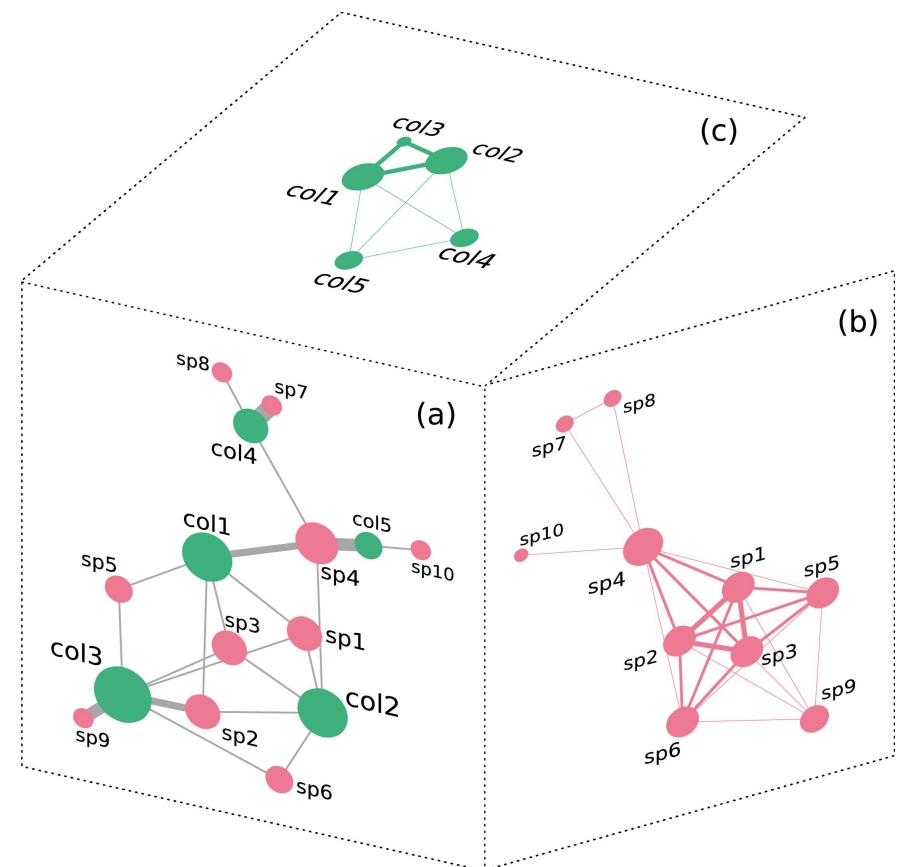
- Perspectiva centrada em espécies
- Vértices de coletores são omitidos
- Conexões se formam entre espécies que foram registradas por pelo menos um coletor em comum

Projeção sobre S_{col}

- Perspectiva centrada em coletores
- Vértices de espécies são omitidos
- Conexões se formam entre coletores que registraram ao menos uma espécie em comum

Regras de pesos

Ajudam a destacar relações mais relevantes
dentre muitas pouco relevantes



Redes Espécie-Coletor (SCN) Agregação taxonômica

Objetivo

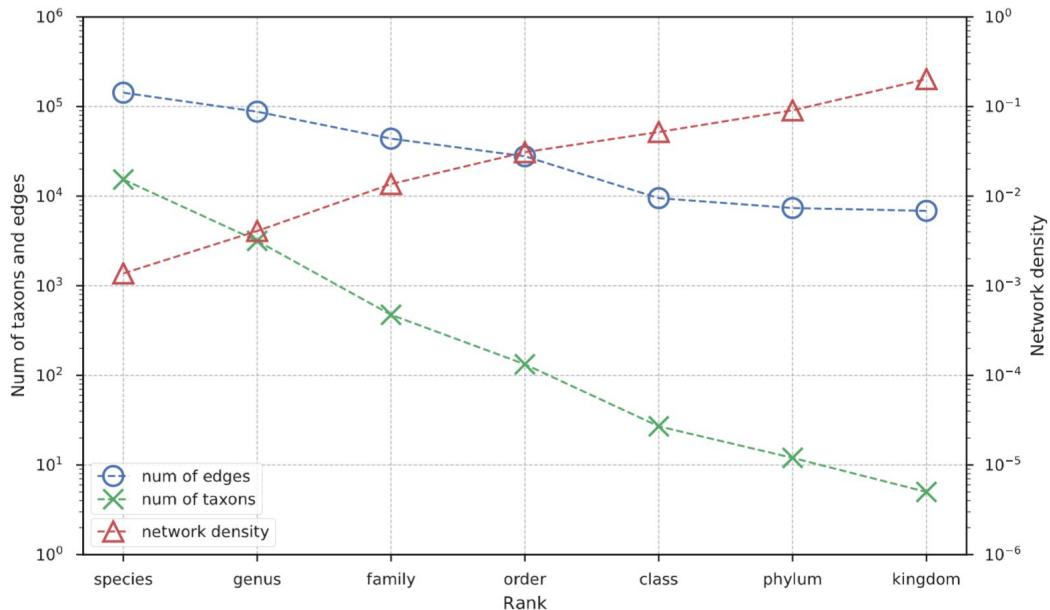
Simplificar uma SCN agrupando espécies em níveis taxonômicos mais altos (ex. família)

Número de vértices em S_{sp} diminui, mas em S_{col} se mantém.

Rotina

- i. Obter um agrupamento de espécies usando uma hierarquia taxonômica;
- ii. Obter o *quorum* para cada espécie;
- iii. Somar os vetores *quorum* para todas as espécies em cada grupo;
- iv. Construir uma nova SCN agregada

A **resolução taxonômica** da SCN é o nível taxonômico no qual a agregação foi obtida



Número de táxons (vértices S_{sp}), arestas e densidade em agregações da SCN (Herbário da UnB) em níveis taxonômicos sucessivos.

Redes Espécie-Coletor (SCN)

Projeções: Regras de ponderação

$w_{(u_s, u_t)} = \sum_{j=1}^m \delta(\sigma_{u_s}^{(j)}, \sigma_{u_t}^{(j)}),$ for the projection onto $S_{col};$

$w_{(v_s, v_t)} = \sum_{i=1}^n \delta(\iota_{v_s}^{(i)}, \iota_{v_t}^{(i)}),$ for the projection onto $S_{sp},$

Ponderação simples

- $(u, v) = 1$ se $u \neq 0$ e $v \neq 0$
- $(u, v) = 0$ caso contrário

Limitação

Pesos somente refletem o número de espécies distintas compartilhadas por dois coletores (ou o número de coletores distintos compartilhados por duas espécies).

O número de recorrências de cada link espécie-coletor na SCN é ignorado

Ponderação média

- $(u, v) = (u+v)/2$ se $u \neq 0$ e $v \neq 0$
- $(u, v) = 0$ caso contrário

Limitação

Coletores com muitos registros de espécies tendem a ser muito mais fortemente ligados entre si;

Pesos sem limite superior, dificultando a filtragem;

Redes Espécie-Coletor (SCN)

Projeções: Regras de ponderação

Ponderação baseada em similaridade

$$sim(\sigma_{u_s}, \sigma_{u_t}) \equiv \cos \theta_{u_s, u_t} = \frac{\sigma_{u_s} \cdot \sigma_{u_t}}{\|\sigma_{u_s}\|_2 \|\sigma_{u_t}\|_2},$$
$$sim(\iota_{v_s}, \iota_{v_t}) \equiv \cos \theta_{v_s, v_t} = \frac{\iota_{v_s} \cdot \iota_{v_t}}{\|\iota_{v_s}\|_2 \|\iota_{v_t}\|_2}.$$

- Função: **Similaridade de cossenos**
- Peso da aresta é dado pela similaridade de *species bags* ou *quorum* de coletores ou espécies
- Valores de pesos limitados no intervalo $[0, 1]$

O Herbário UB

Caracterização Geral

Taxonômica

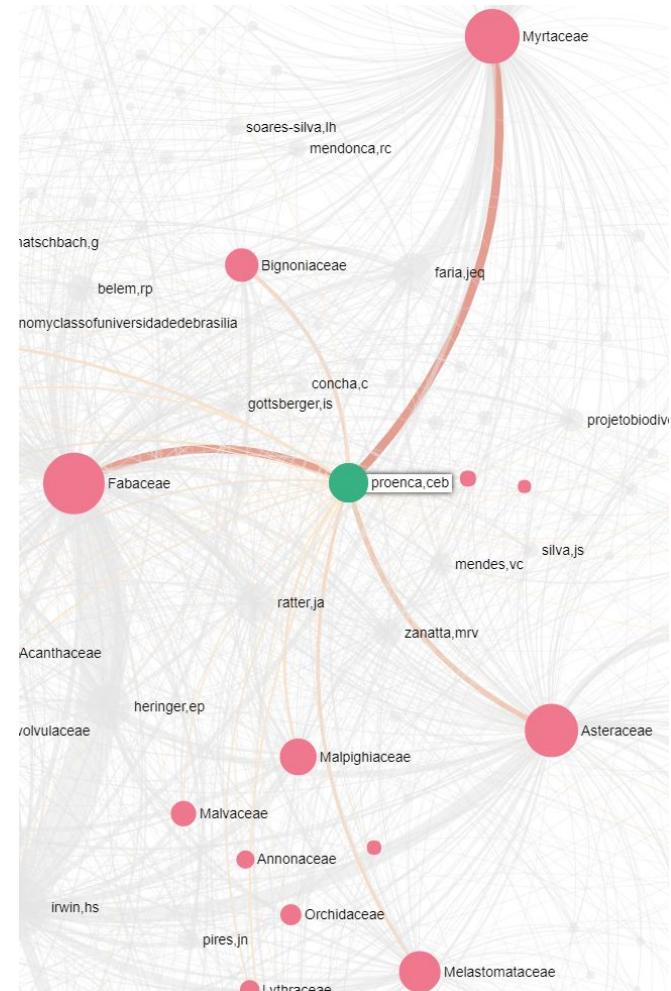
- Família *Fabaceae* é a mais representativa, seguida por *Myrtaceae*, *Asteraceae*, *Rubiaceae*

	num of taxa	top-4 taxa	num of records	% of records
family	507	Fabaceae	19254	10.39
		Myrtaceae	12833	6.93
		Asteraceae	11271	6.08
		Rubiaceae	9447	5.10
genus	3374	Myrcia	4654	2.51
		Eugenia	3750	2.02
		Mimosa	2992	1.61
		Miconia	2402	1.30

O Herbário UB Rede Espécie-Coletor (SCN) Coletores importantes

Carolyn E. B. Proença (*proenca,ceb*)

- Professora e Pesquisadora na UnB, desde 1992
- Perfil levemente generalista, com ênfase em *Myrtaceae*, *Fabaceae*, *Asteraceae*, *Melastomataceae*, *Bignoniaceae*



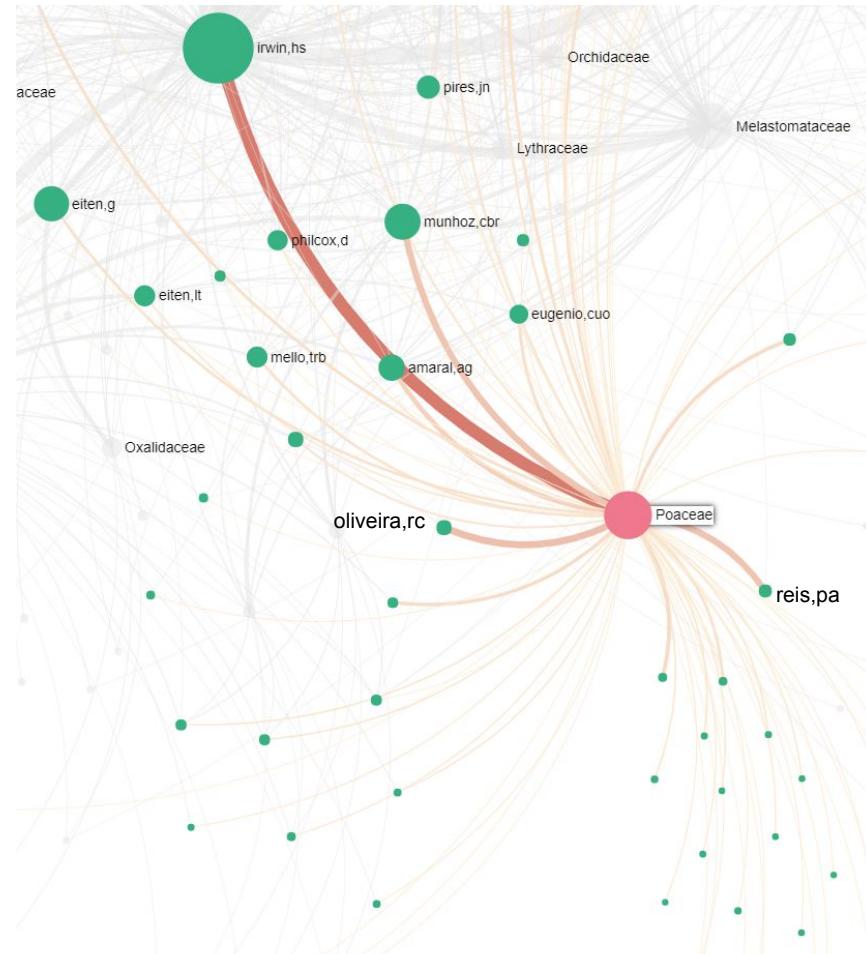
https://lncc-netsci.github.io/pedrocs/networks/ub_scn/

O Herbário UB Rede Espécie-Coletor (SCN) Famílias importantes



Poaceae

- Família diversa, composta por espécies de capins
- Coletada por um conjunto mais **especializado** de coletores (embora também por alguns **generalistas**):
 - Howard Irwin
 - Regina Célia Oliveira
 - Priscila Alves dos Reis
 - Cássia B. R. Munhoz



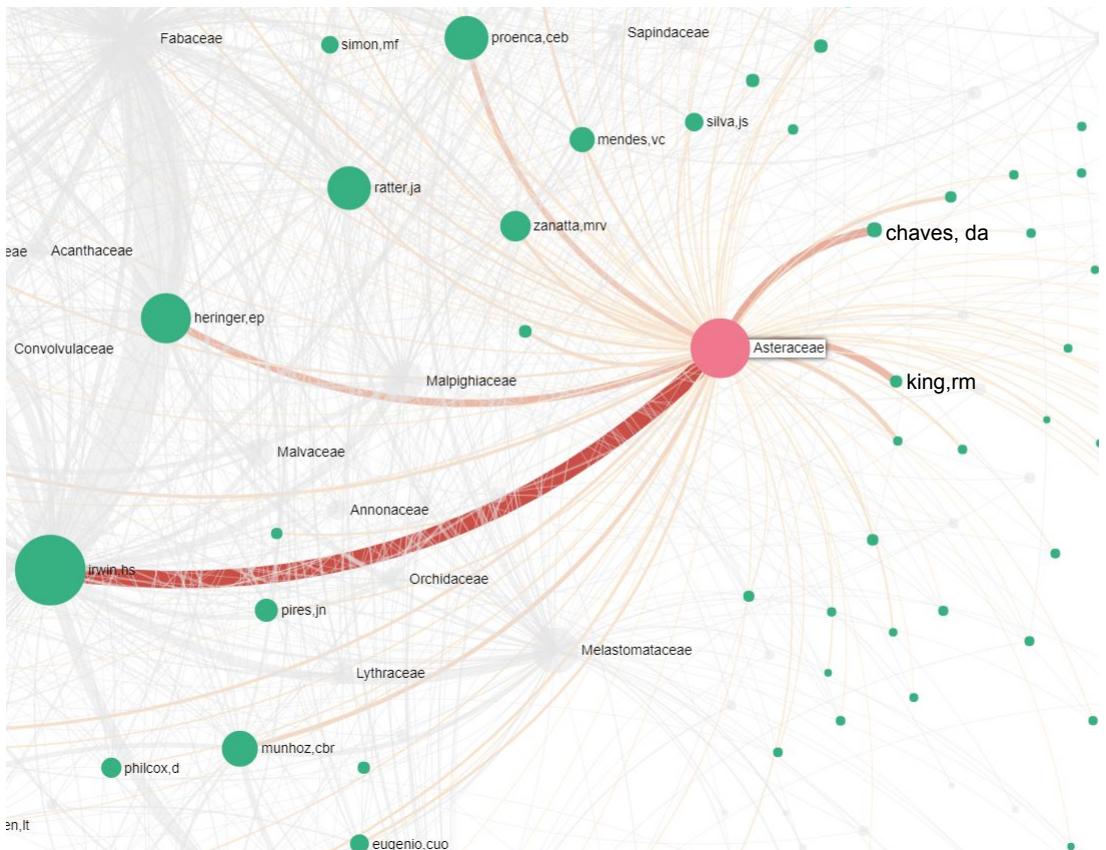
https://lncc-netsci.github.io/pedrocs/networks/ub_scn/

O Herbário UB Rede Espécie-Coletor (SCN) Famílias importantes



Asteraceae

- Família bastante diversa de plantas herbáceas
- Coletada tanto por **especialistas** quanto **generalistas**:
 - Howard Irwin
 - Ezechias P. Heringer
 - Carolyn E. B. Proença
 - Daniel A. Chaves
 - Robert Merrill King



O Herbário UB Rede Colaborativa de Coletores (CWN)

Tamanhos das equipes

Quantidade de registros por tamanho de equipe decresce exponencialmente ($b_0 = 0.16$, $b_1 = -0.38$):

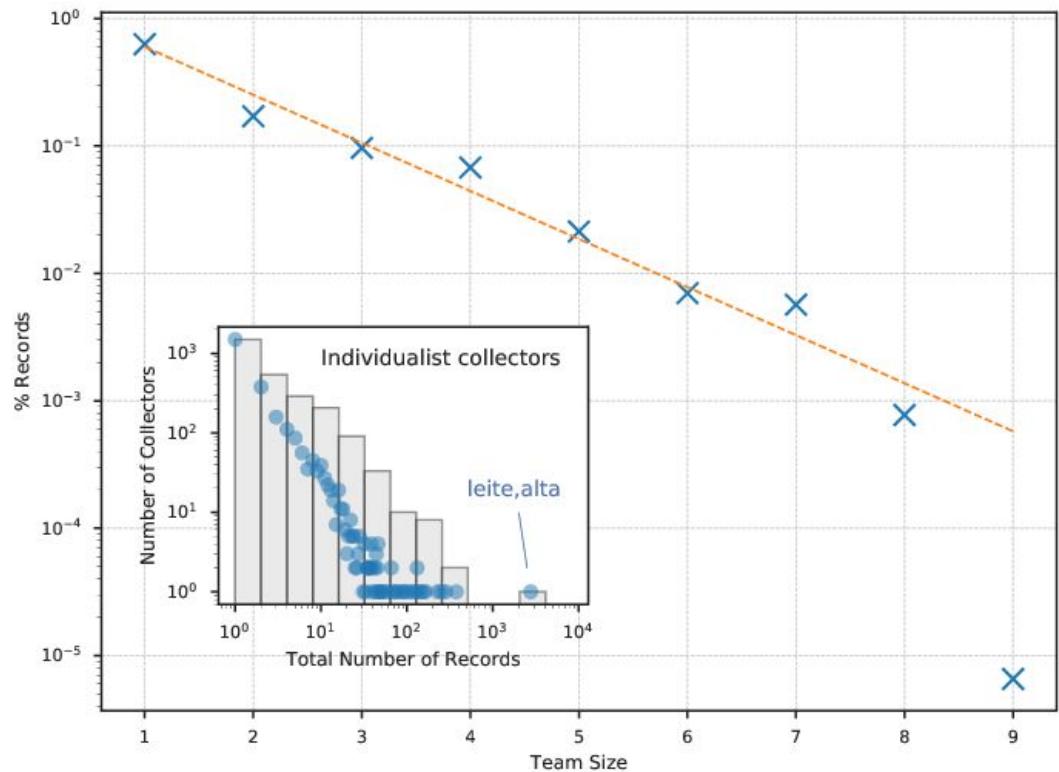
$$f(x) = 10^{\beta_0 + \beta_1 x}$$

Coletores individualistas

Muitos coletores com poucos registros, poucos coletores com muitos registros

Ana Lúcia T. Leite é a coletrora individualista com mais registros (2757 registros)

Todos os outros têm menos que 400 registros



O Herbário UB

Rede Colaborativa de Coletores (CWN)

Top-20 coletores (k_w)

collector	num of records	% collaborative	k	k_w
irwin,hs	18065	34.1	39	5696.0
proenca,ceb	4803	88.3	218	4203.7
faria,jeq	4687	82.9	117	3881.0
souza,rr	3885	99.3	37	3835.5
santos,rrb	3587	94.5	41	3382.3
munhoz,cbr	3191	82.8	109	2493.2
zanatta,mrv	2364	95.8	50	2264.0
castelobranco,cw	2256	100.0	1	2256.0
grando,jv	2256	100.0	1	2256.0
eiten,g	3046	61.4	33	1865.5
amaral,ag	1825	99.5	37	1815.0
projeto biodiversidade bp	1780	100.0	19	1780.0
mendes,vc	1696	99.9	89	1695.0
fonseca,sf	1610	99.9	18	1607.0
camara,peas	2076	79.8	47	1486.0
harley,rm	2564	66.0	90	1455.8
carvalhosilva,m	1635	98.5	58	1436.0
eiten,lt	1262	99.6	14	1256.5
mello,trb	1247	100.0	21	1247.0
soares,aer	1557	73.0	29	1135.0

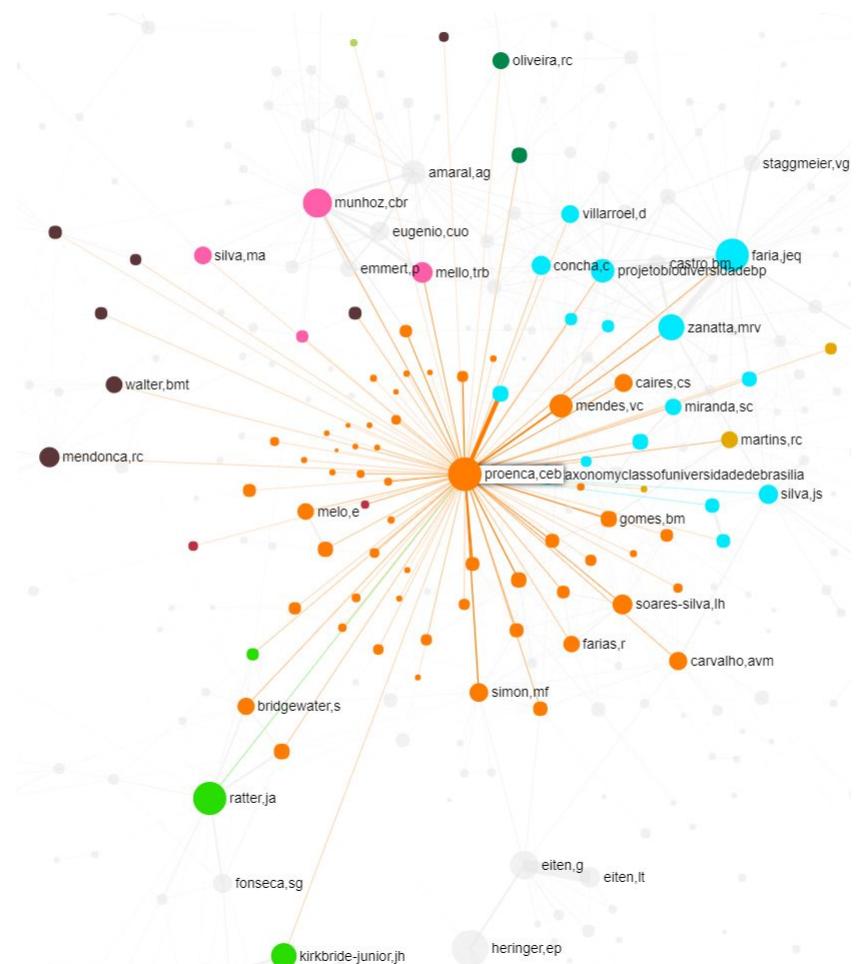
O Herbário UB

Rede Colaborativa de Coletores (CWN)

Comunidades de colaboração

Carolyn E. B. Proença (*proenca,ceb*)

- Conecta muitos coletores de comunidades diferentes.
- É o vértice mais central na rede: Importante intermediário entre vários coletores



O Herbário UB

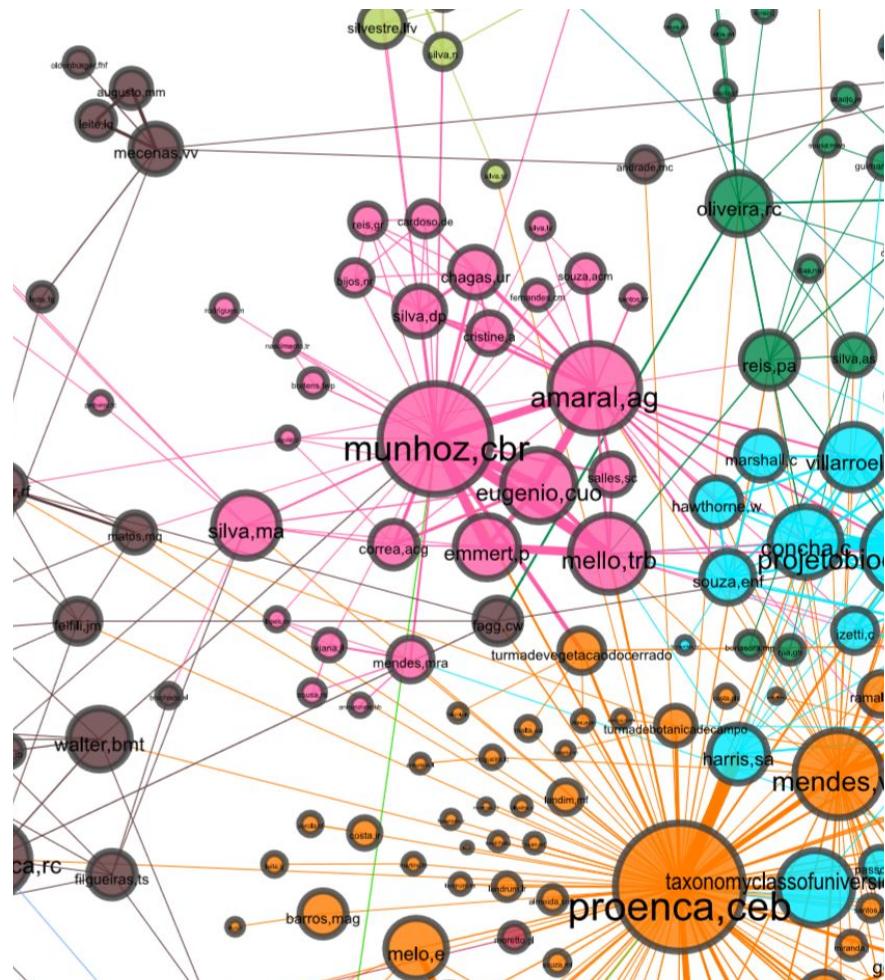
Rede Colaborativa de Coletores (CWN)

Comunidades de colaboração

Interesses Similares

**Cássia Munhoz, Aryanne Amaral,
Chesterton U. Eugênio**

- Coletores de plantas herbáceas
-



O Herbário UB

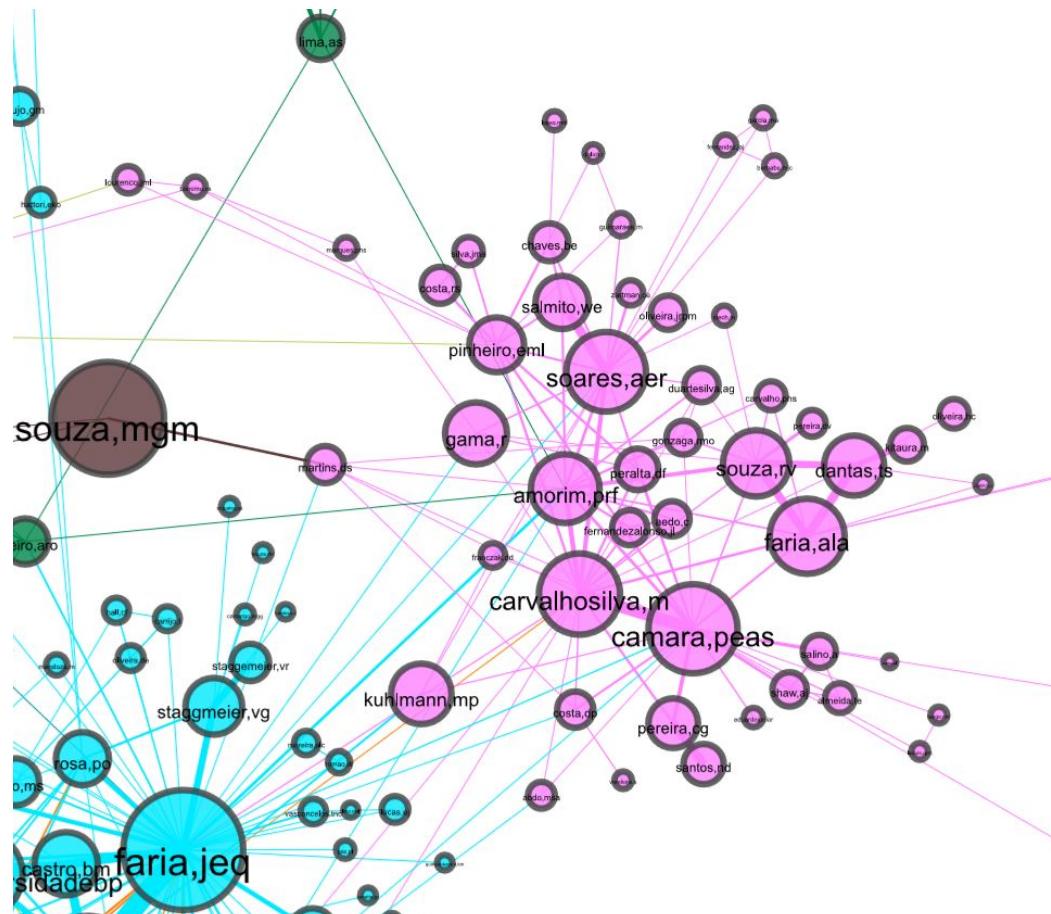
Rede Colaborativa de Coletores (CWN)

Comunidades de colaboração

Interesses Similares

Laboratório de Criptógamas

- Vértices também formam uma comunidade na SCN
 - Coletores interessados em briófitas também colaboram em campo
 - Principais coletores:
 - Paulo E. Câmara
 - Micheline Carvalho Silva
 - Maria das Graças Machado

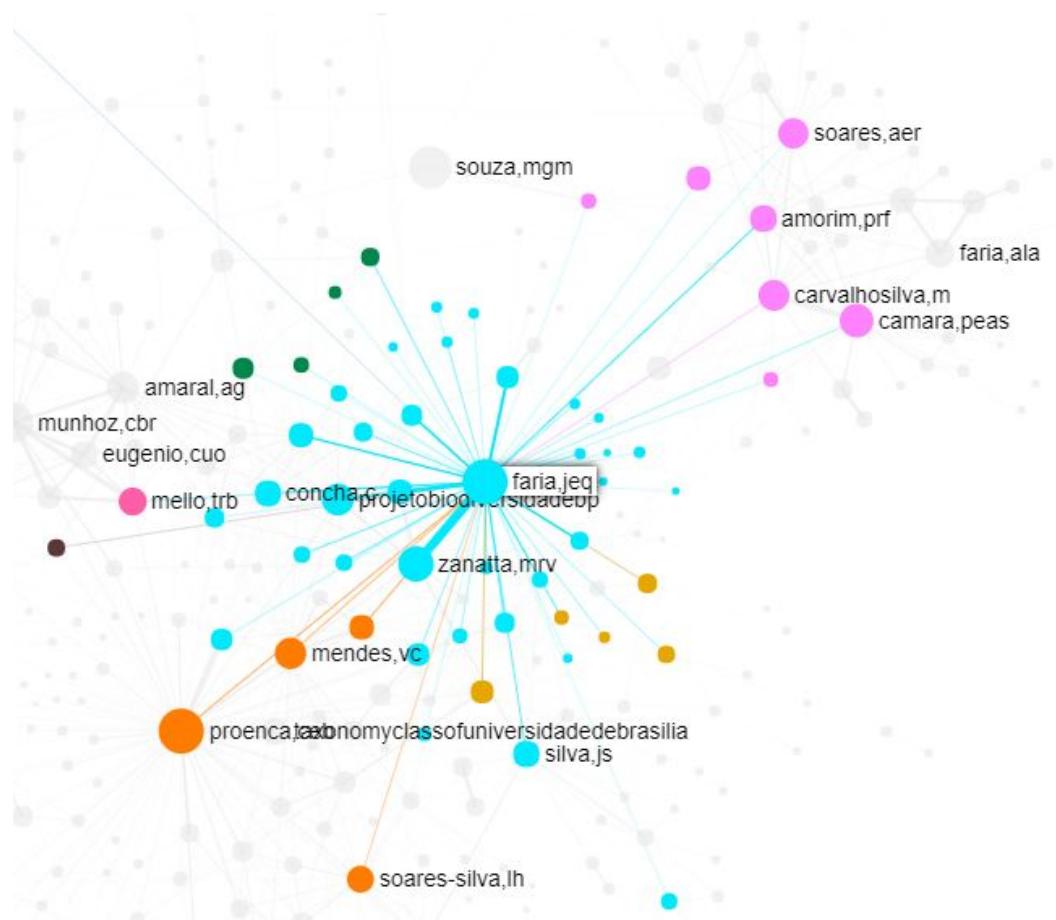


O Herbário UB Rede Colaborativa de Coletores (CWN)

Comunidades de colaboração

Jair E. Q. Faria

- Um importante intermediário entre coletores no grupo de briófitas com o centro da rede



O Herbário UB

Rede Colaborativa de Coletores (CWN)

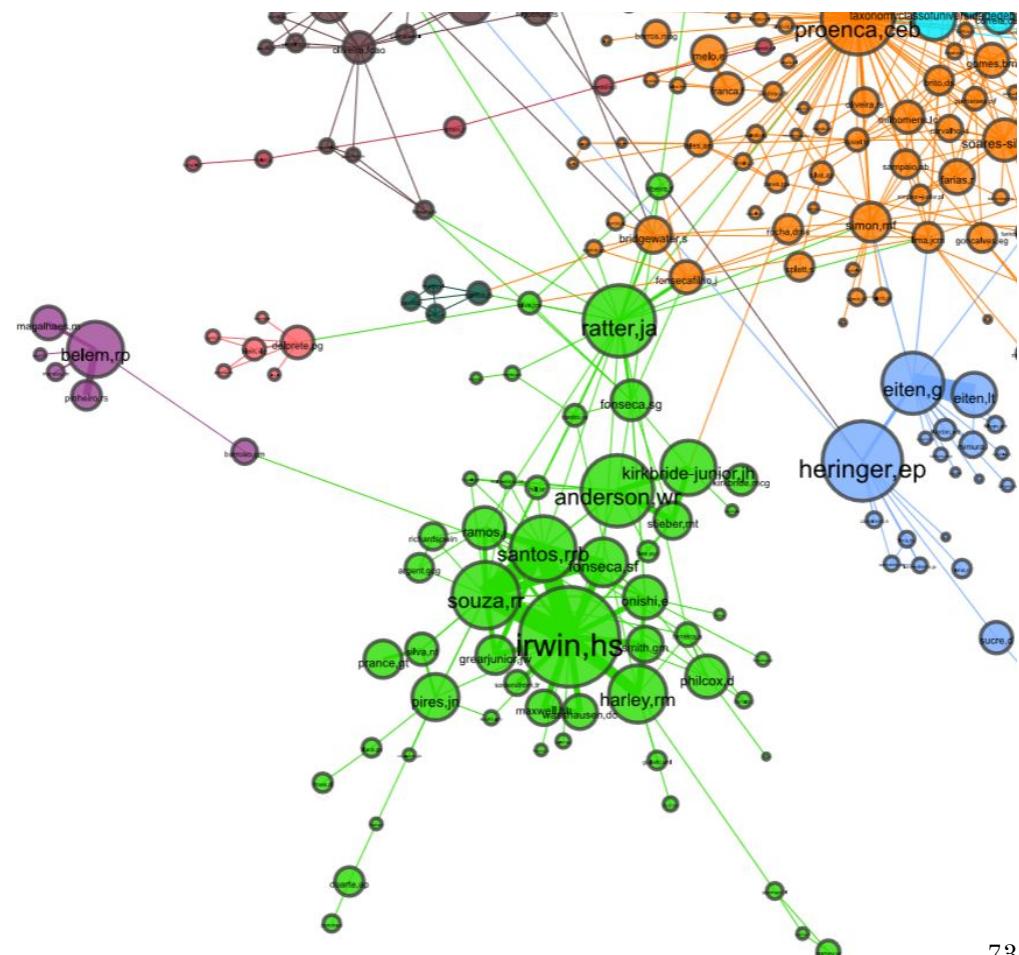
Comunidades de colaboração

Separação temporal

Expedição Planalto Central

Howard S. Irwin (*irwin,hs*)

- Howard S. Irwin, William R. Anderson, R. M. Harley
 - Perfil levemente generalista, com ênfase em *Myrtaceae*, *Fabaceae*, *Asteraceae*, *Melastomataceae*, *Bignoniaceae*



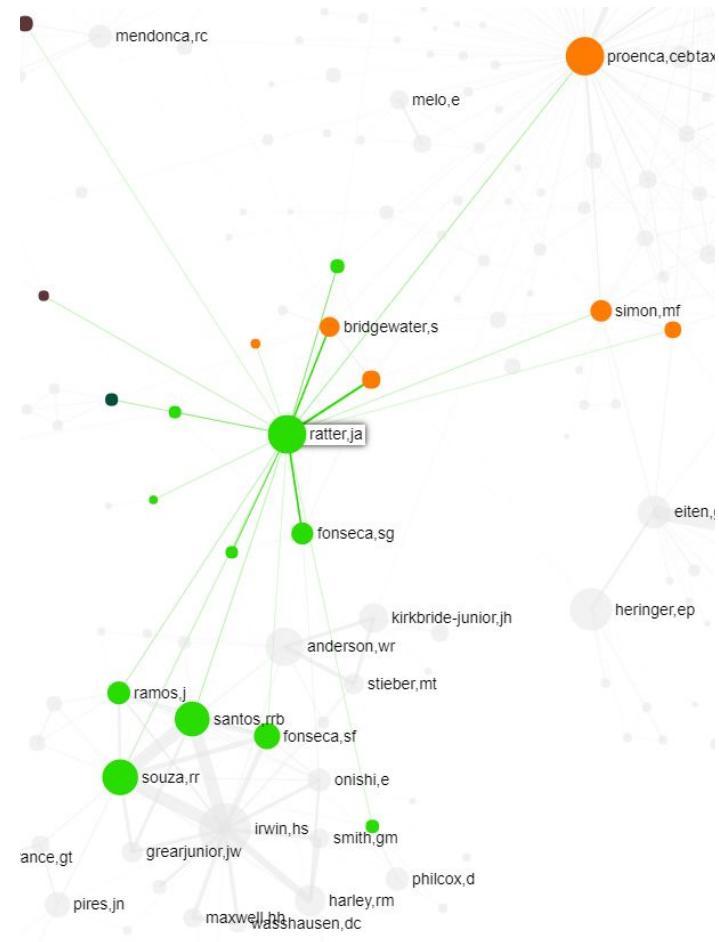
O Herbário UB

Rede Colaborativa de Coletores (CWN)

Comunidades de colaboração

James A. Ratter (*ratter,ja*)

- Um importante intermediário “temporal”: coletou com duas gerações diferentes



O Herbário UB Rede Colaborativa de Coletores (CWN)

Comunidades de colaboração

Separação temporal

Ezechias P. Heringer

- blablabá

George Eiten & Liene Eiten

