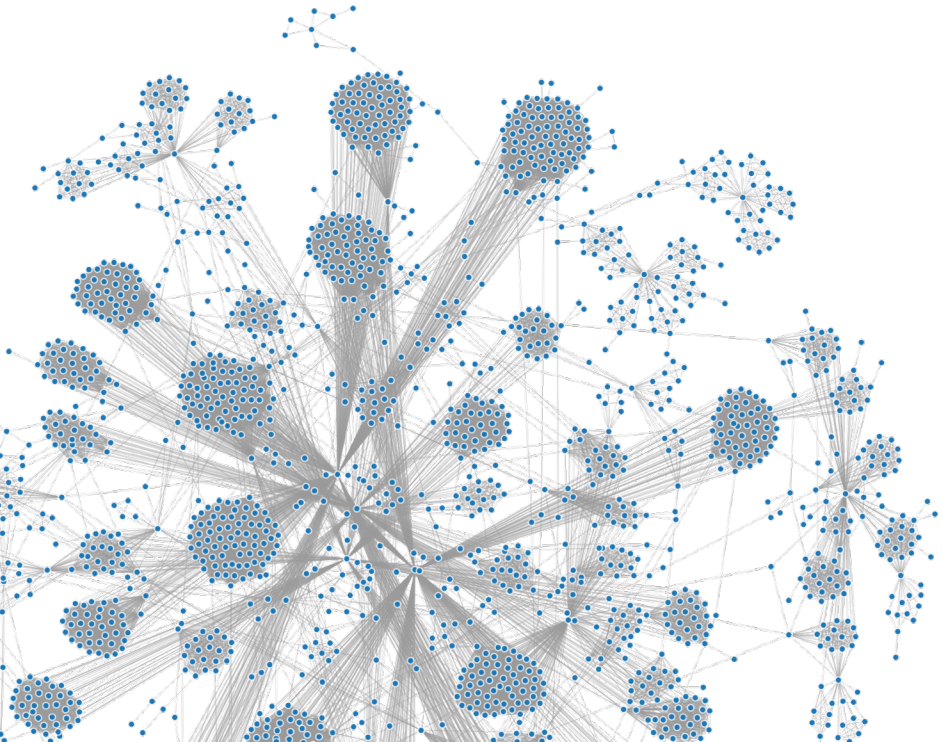
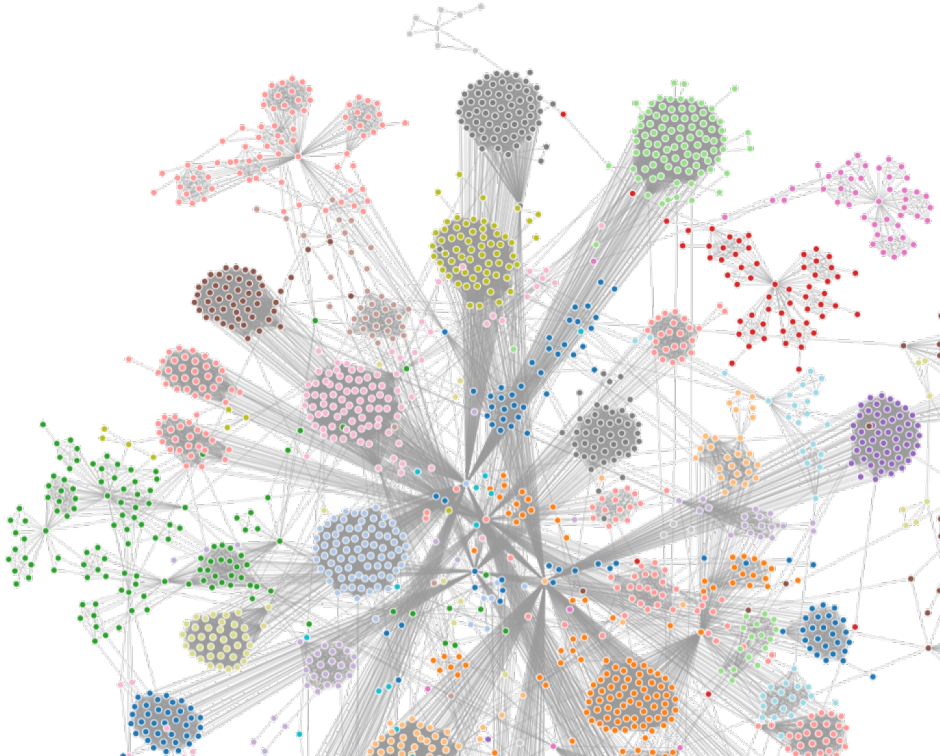
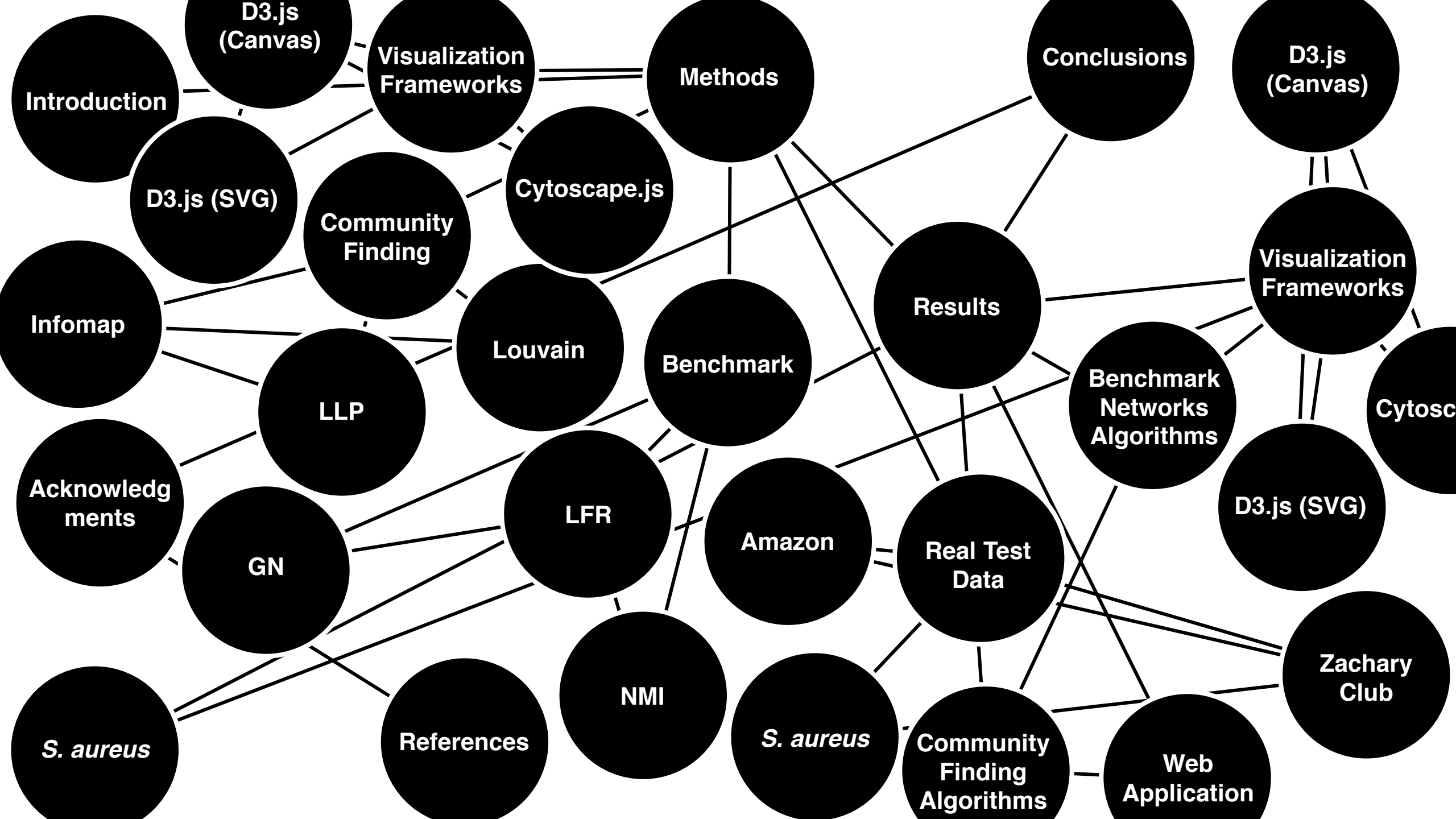
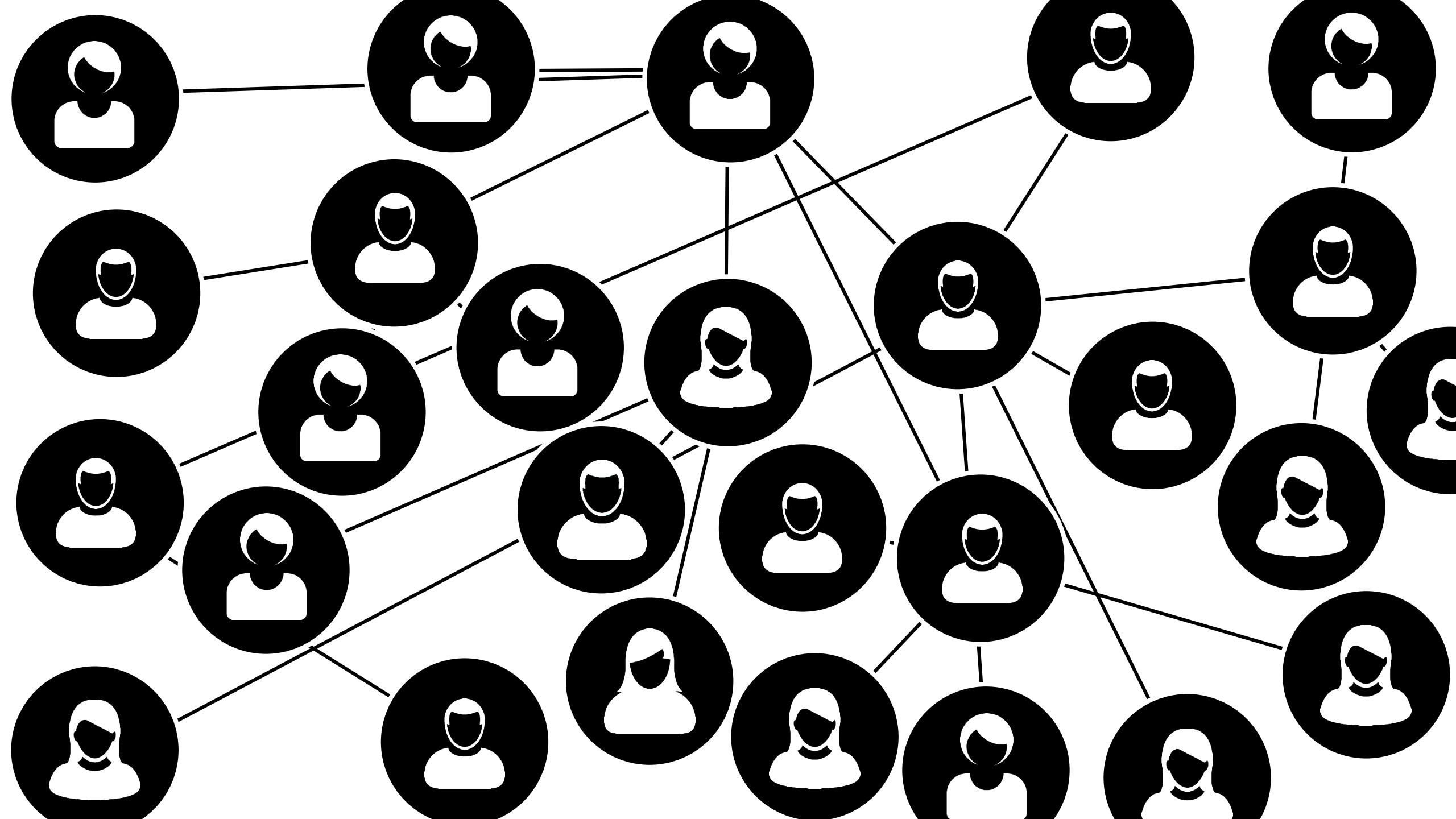
Community Finding with Applications on Phylogenetic Networks

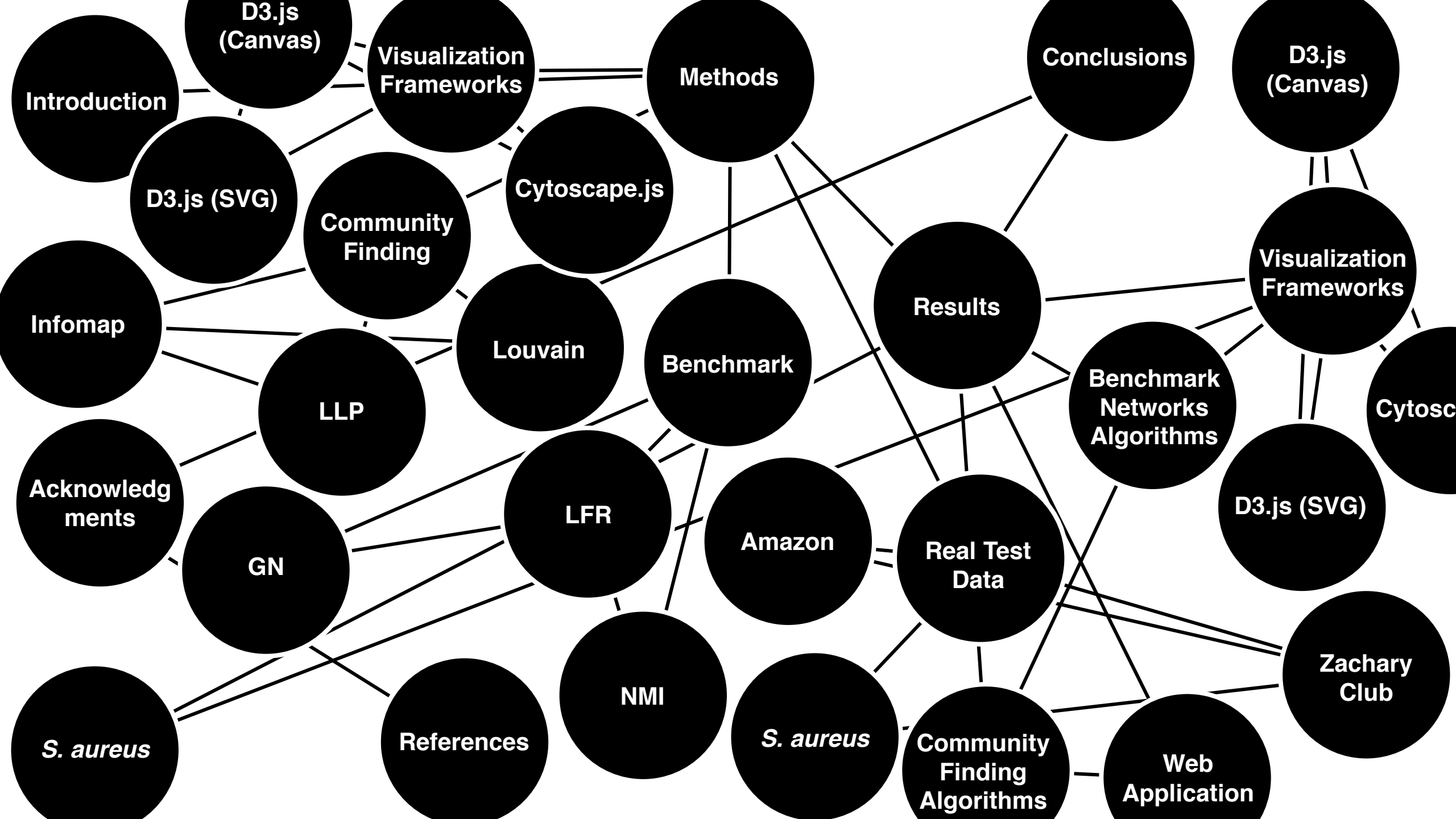
Luís Rita

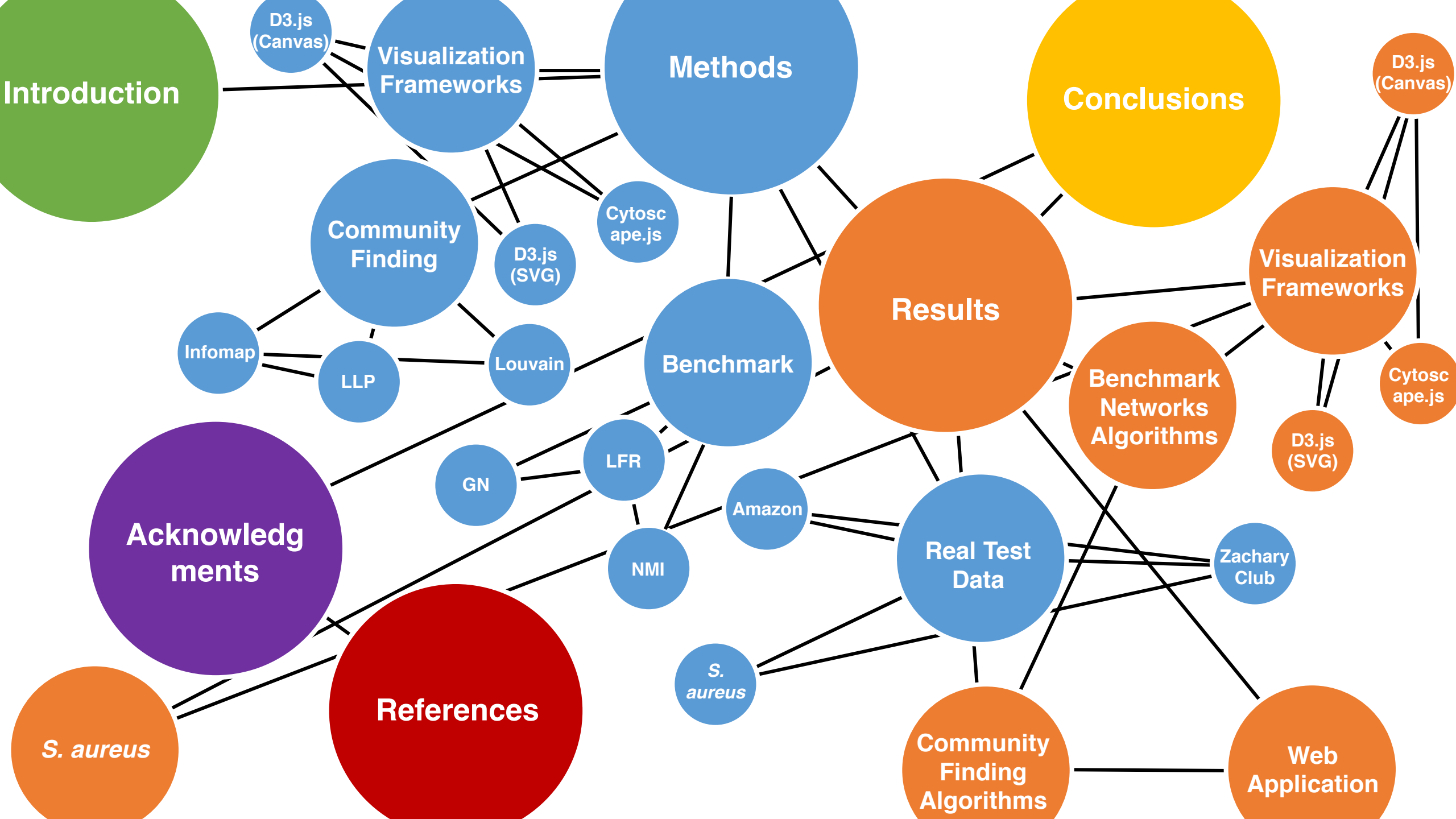





Introduction
D3.js (Canvas)
Visualization Frameworks
Methods
Conclusions
D3.js (Canvas)
D3.js (SVG)
Community Finding
Cytoscape.js
Infomap
Louvain
Benchmark
Results
Visualization Frameworks
LLP
Benchmark Networks Algorithms
Cytosc
Acknowledgments
GN
LFR
Amazon
Real Test Data
D3.js (SVG)
Zachary Club
NMI
S. aureus
Community Finding Algorithms
Web Application
S. aureus
References




Introduction
D3.js (Canvas)
Visualization Frameworks
Methods
Conclusions
D3.js (Canvas)
D3.js (SVG)
Community Finding
Cytoscape.js
Infomap
Louvain
Benchmark
Results
Visualization Frameworks
LLP
Benchmark Networks Algorithms
Cytosc
Acknowledgments
GN
LFR
Amazon
Real Test Data
D3.js (SVG)
Zachary Club
NMI
S. aureus
Community Finding Algorithms
Web Application
S. aureus
References


Introduction
D3.js (Canvas)
Visualization Frameworks
Methods
Conclusions
D3.js (Canvas)
Community Finding
D3.js (SVG)
Cytoscape.js
Results
Visualization Frameworks
Infomap
LLP
Louvain
Benchmark
Benchmark Networks Algorithms
Cytoscape.js
D3.js (SVG)
GN
LFR
Amazon
Real Test Data
Zachary Club
Acknowledgments
NMI
S. aureus
References
S. aureus
Community Finding Algorithms
Web Application

Introduction

“The world’s most valuable resource is no longer oil, but data.”

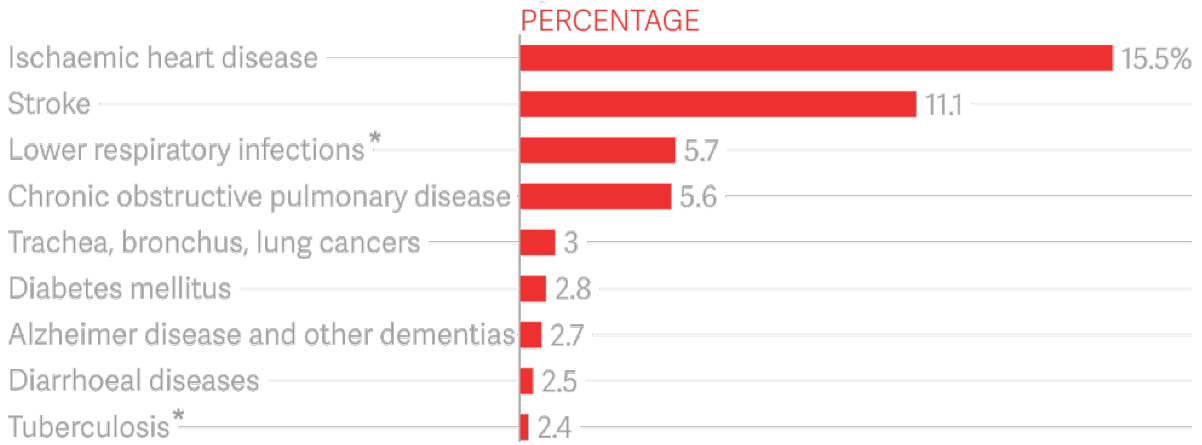
The Economist

“The rapid advances in NGS technologies and capabilities has proven and further promises to be a game changer in diagnostic microbiology (…)”

BDQ Journal (Elsevier)

“Ten threats to global health in 2019 – Antimicrobial Resistance”

World Health Organization



Methods

Community Finding

Louvain

Compute $Mod_{new} = Mod(M)$
repeat
  $Mod = Mod_{new}$
  Randomize the order of vertices.
  for all $i \in V^k$ **do**
    $best_{com} = M_i^k; best_{increase} = 0$
    for all $M' \in C^k$ **do**
      $M_i^k = M_i^k \backslash \{i\}$
      $\Sigma_{tot}{}^{M_i^k} = \sum_{\alpha \in M_i^k} k_\alpha - k_i; \Sigma_{tot}{}^{M'^k_i} = \sum_{\alpha \in M'^k_i} k_\alpha + k_i;$
      $\Sigma_{in} = \Sigma_{tot}{}^{M'^k_i} - \sum k_{i,j}, (i,j) \in E^k, i \in C_i^k \; and \; j \notin C_i^k$
      $k_{i,in} = \sum_{\alpha \in M'^k_i} k_{i,\alpha}$
      If $\delta Mod_{M_i^k \to M'^k_i} > best_{increase}$ then
        $best_{increase} = \delta Mod_{M_i^k \to M'^k_i}$

$$M = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

$$\Delta Mm = k_{i,in} - \frac{\sum_{tot} k_i}{2m}$$

Infomap

Compute $L_{new} = L(M)$
repeat
  $L = L_{new}$
  Randomize the order of vertices.
  for all $u \in V^k$ do
    if $M'^k_u = argmin \left(\delta L_{M_u^k \to M'^k_u} \right) < 0$ then
      $M_u^k = M_u^k \backslash \{u\}; M'^k_u = M'^k_u \cup \{u\}$
      $p^{M_u^k} = \sum_{\alpha \in M_u^k} p_\alpha - p_u; p^{M'^k_u} = \sum_{\alpha \in M'^k_u} p_\alpha + p_u$
      update $q^{M_u^k}$; update $q^{M'^k_u}$
    end if
  end for
  Compute $L_{new} = L(M)$
until No vertex movement or $L - L_{new} < \theta$
Compute $L_{new} = L(M)$

$$L(M) = \left(\sum_{m \in M} q_m \right) log \left(\sum_{m \in M} q_m \right) - 2 \sum_{m \in M} q_m \, log(q_m) - \sum_{\alpha \in V} p_\alpha \, log(p_\alpha) + \sum_{m \in M} \left(q_m + \sum_{\alpha \in m} p_\alpha \right) log \left(q_m + \sum_{\alpha \in m} p_\alpha \right)$$

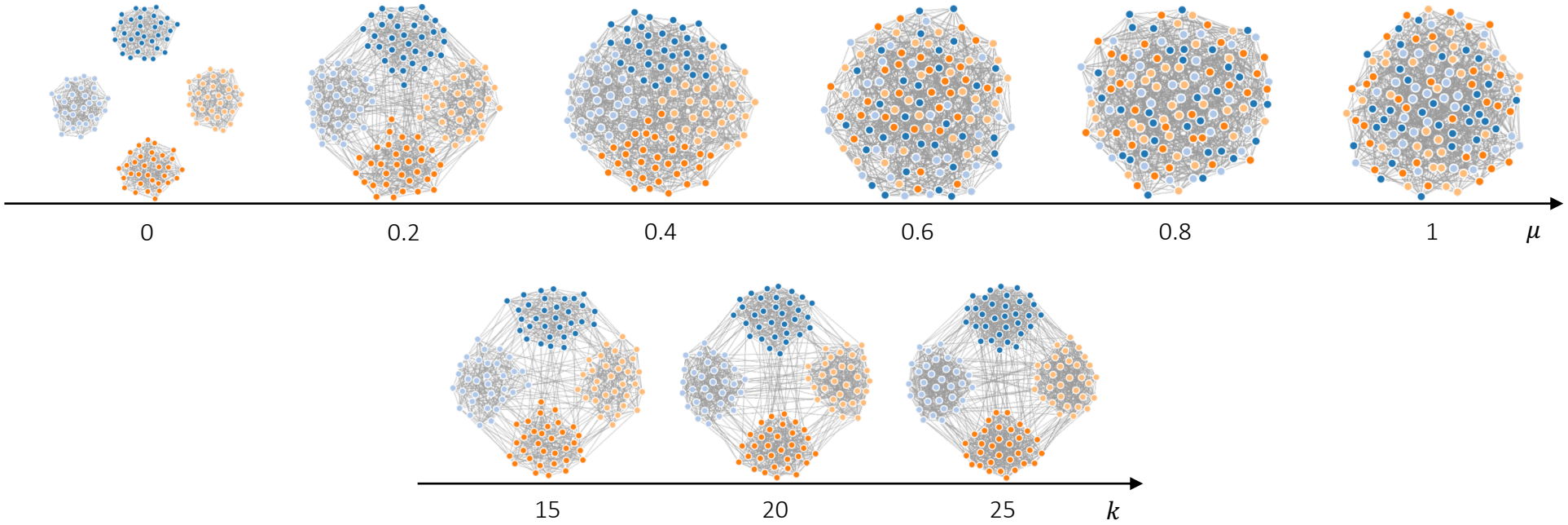
Layered Label Propagation

$k = 0$
for all $i \in V^0$ **do**
  $M_i^0 = \{i\}$
end for
repeat
  Randomize the order of vertices.
  for all $u \in V^k$ **do**
    for all $i \in labels(u)$ **do**
      If $k_i(u) - \gamma(v_i(u) - k_i(u)) > k_{i-1}(u) - \gamma(v_{i-1}(u) - k_{i-1}(u))$ then
        $M_u^k = M_u^k \backslash \{u\}; M'^k_u = M'^k_u \cup \{u\}$
      end if
    end for
  end for
$k = k + 1$
until No vertex movement or $k < max_{iteration}$

$$k_i - \gamma(v_i - k_i)$$

Methods

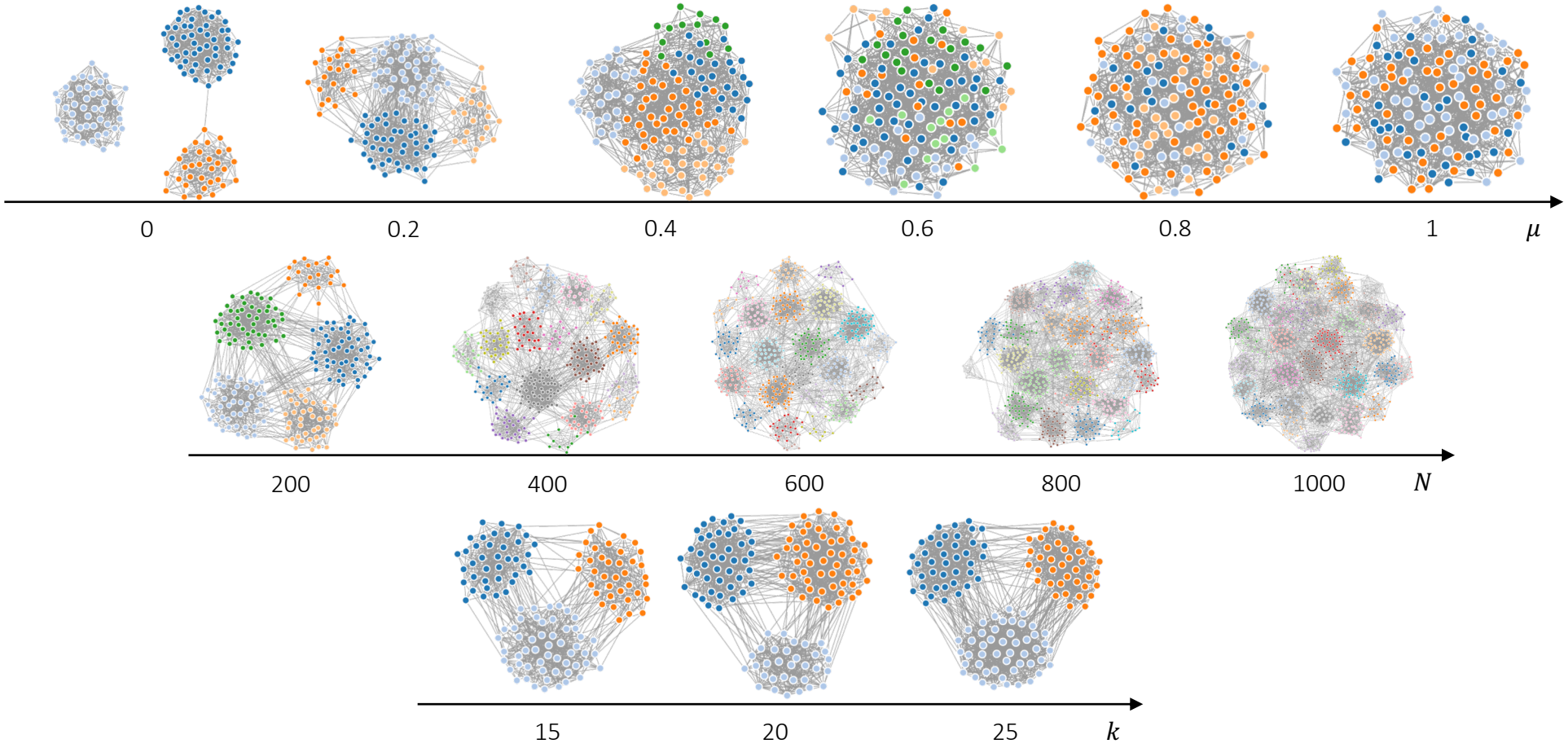
Benchmark Networks



Girvan - Newman

Methods

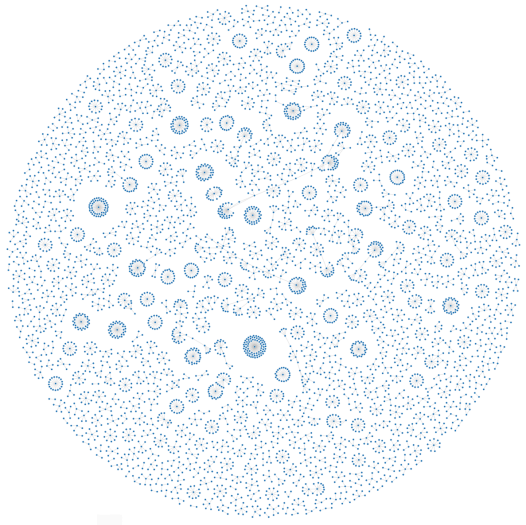
Benchmark Networks



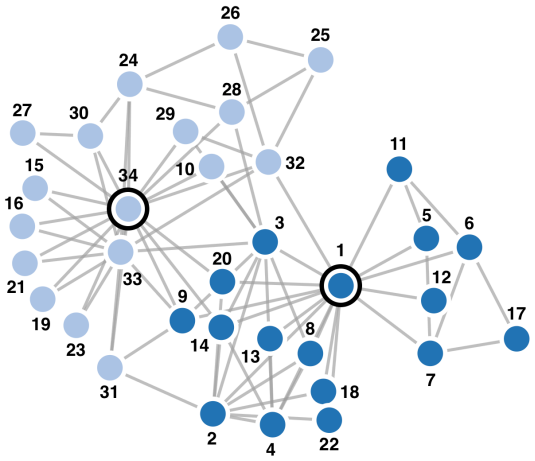
Lancichinetti - Fortunato - Radicchi

Methods

Real Test Data



Amazon

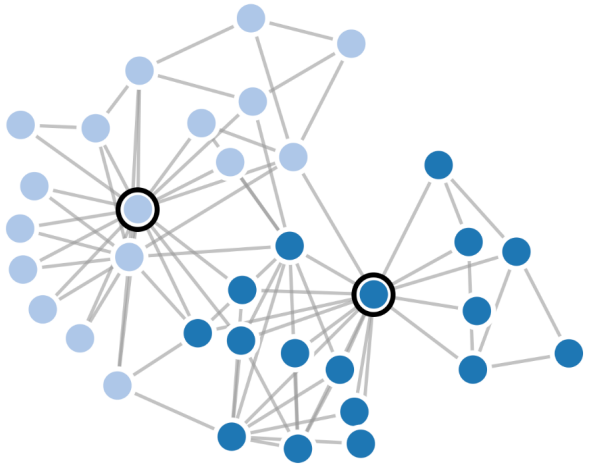


Zachary's Karate Club

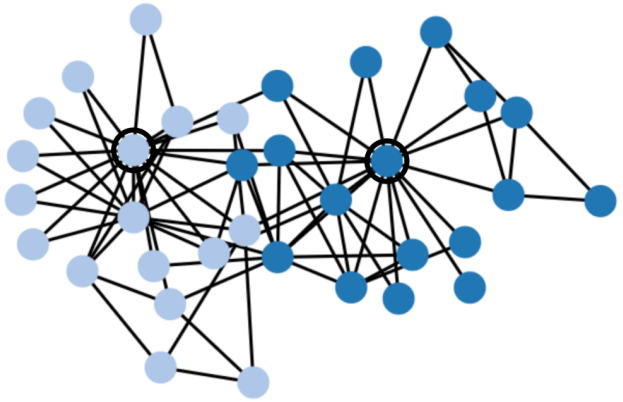
Staphylococcus aureus
MLST SLV Network

Methods

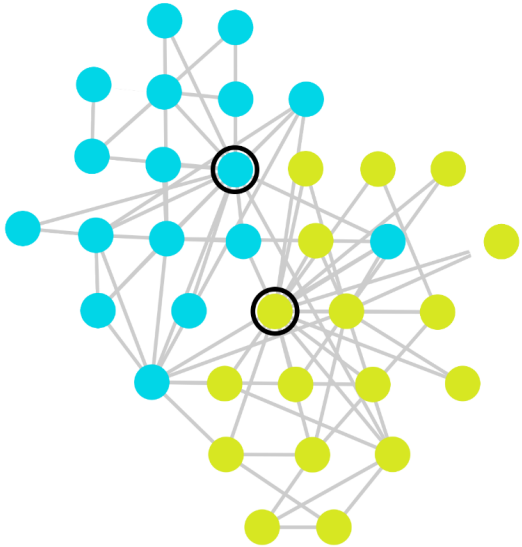
Visualization Frameworks



D3.js & SVG



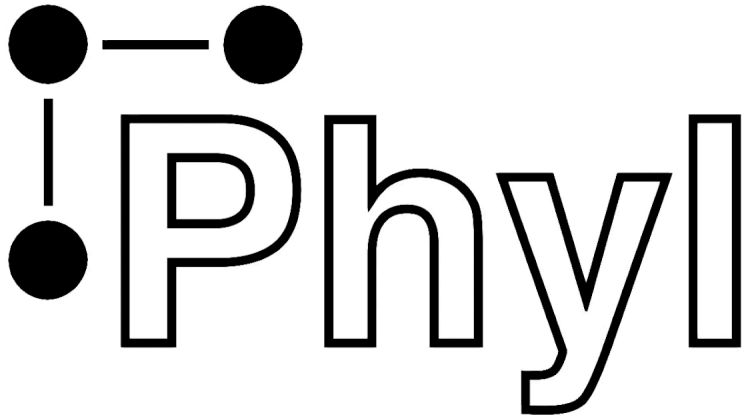
D3.js & Canvas



Cytoscape.js

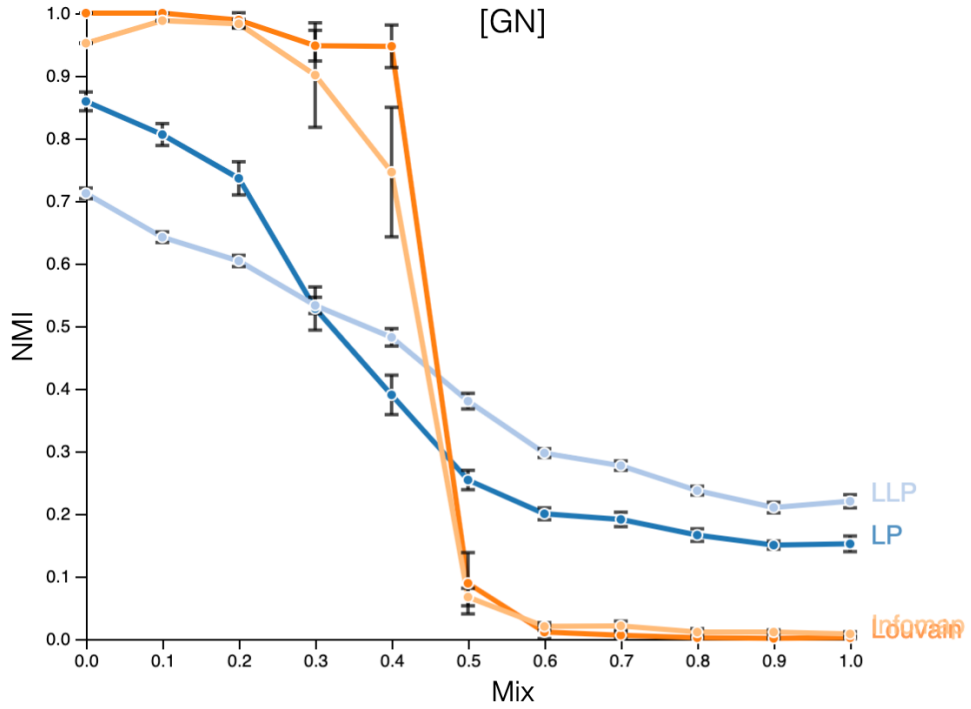
Results

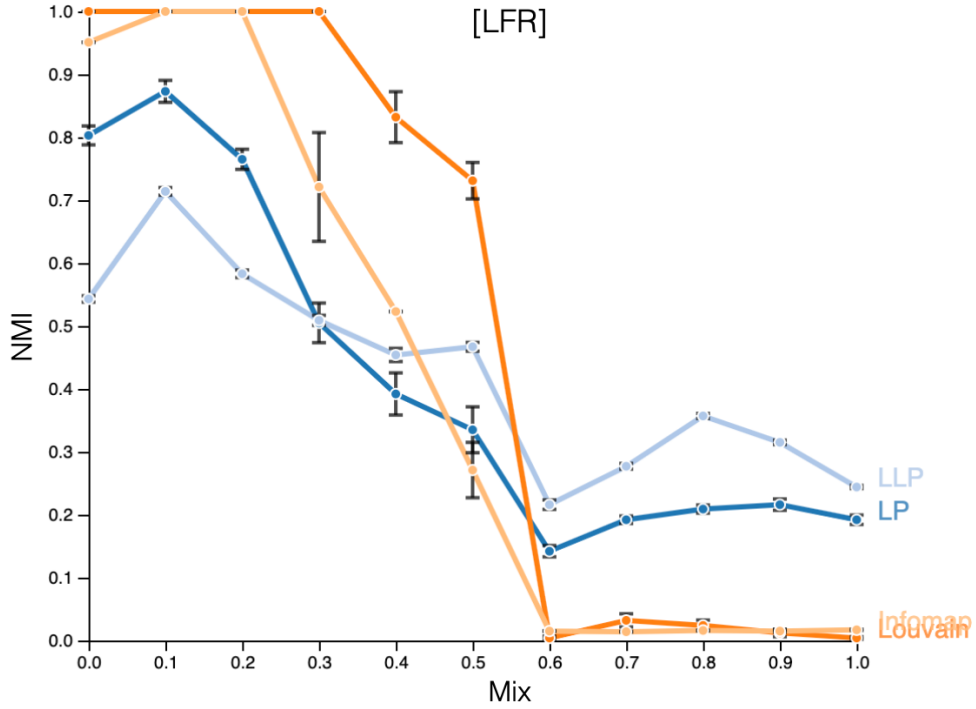
Web Application



Results

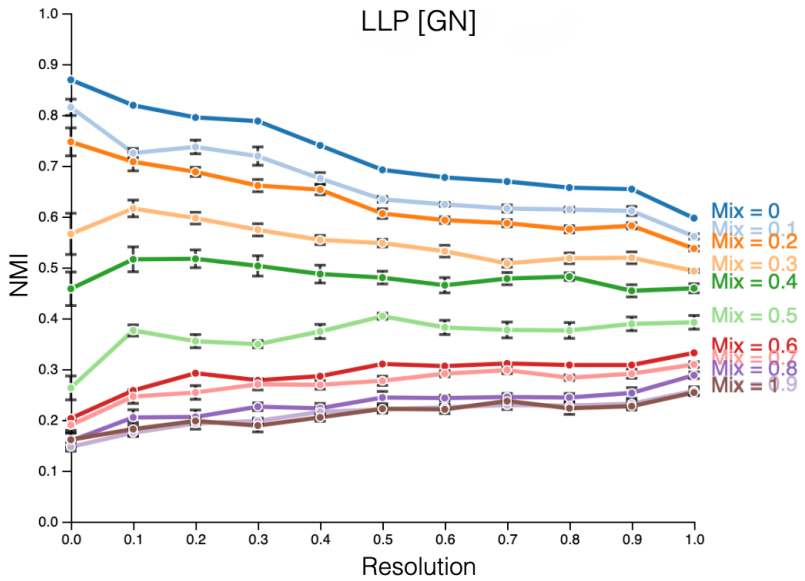
Community Finding Algorithms



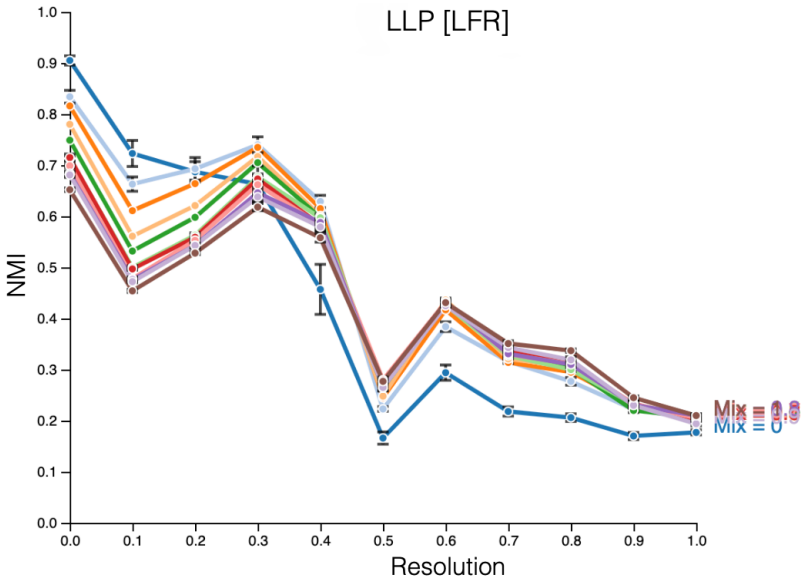


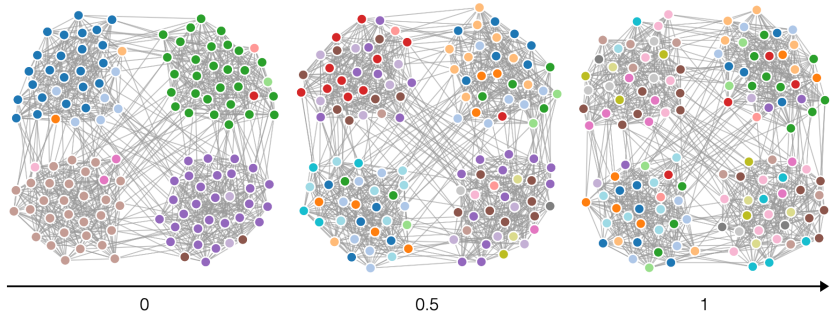
Results

Community Finding Algorithms



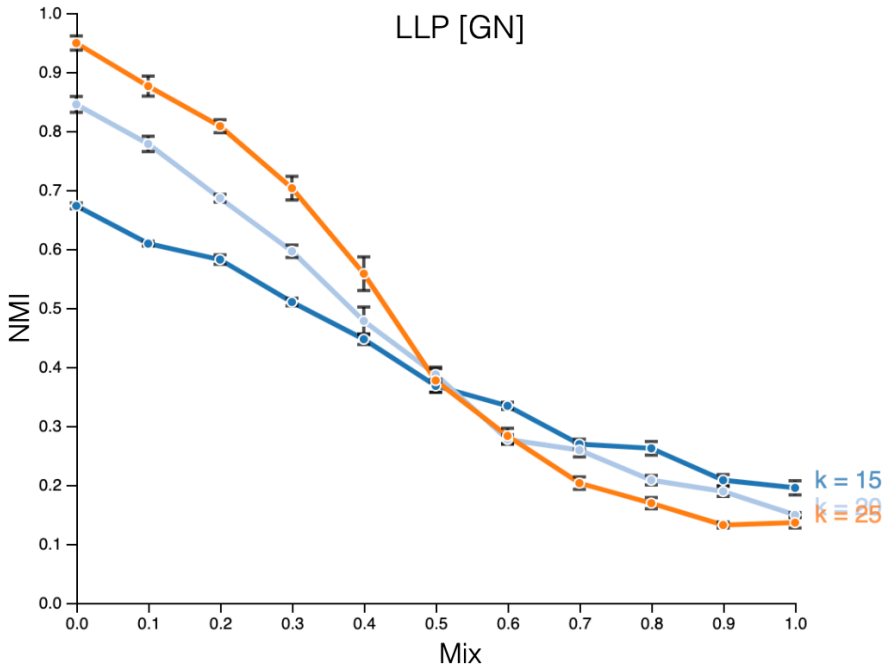
$$k_i - \gamma(v_i - k_i)$$

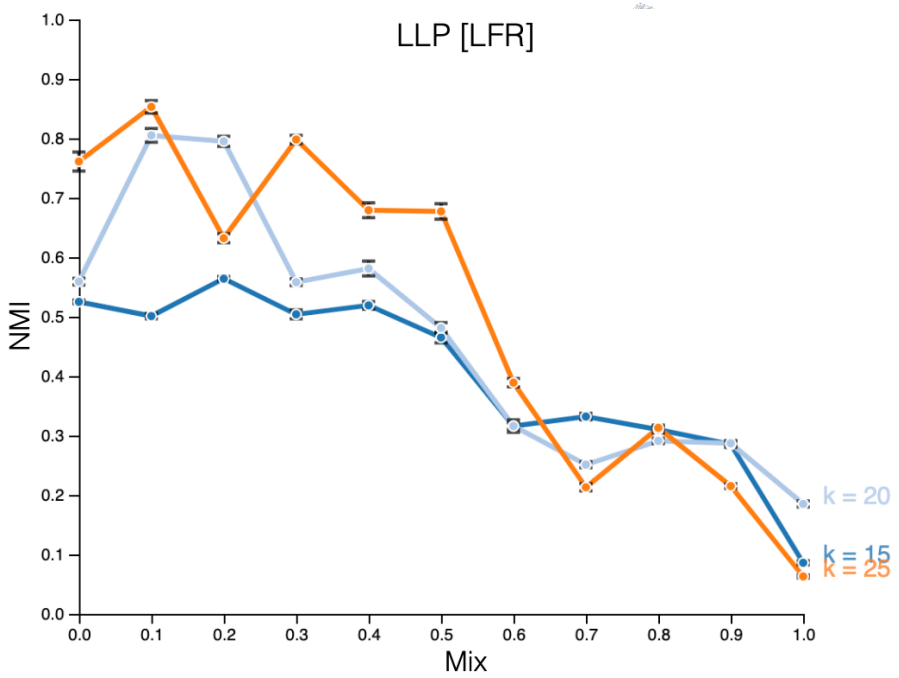




Results

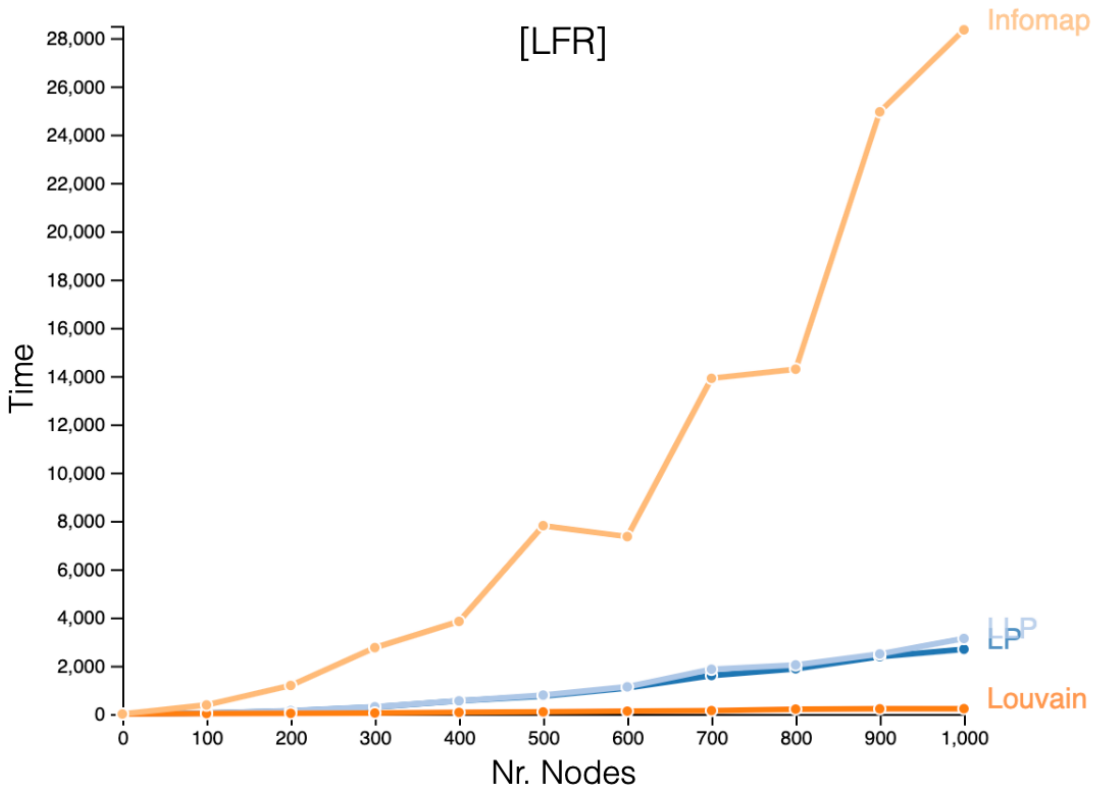
Community Finding Algorithms





Results

Community Finding Algorithms



Algorithm	Time Complexity
Louvain	$O(L)$
Infomap	$O(NlogN)$
LP	$O(L)$
LLP	$O(L)$

$$\Delta Mm = k_{i,in} - \frac{\sum_{tot} k_i}{2m}$$

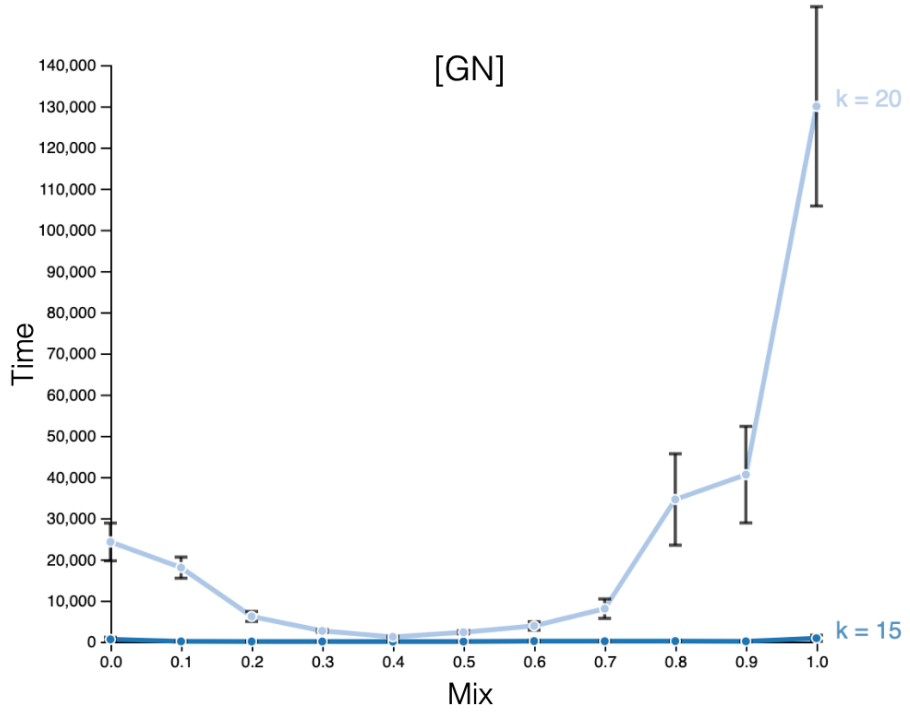
$$= \left(\sum_{m \in M} q_m\right) log\left(\sum_{m \in M} q_m\right) - 2\sum_{m \in M} q_m \, log(q_m) - \sum_{\alpha \in V} p_\alpha \, log(p_\alpha) + \sum_{m \in M}\left(q_m + \sum_{\alpha \in m} p_\alpha\right) log\left(q_m + \sum_{\alpha \in m} p_\alpha\right)$$

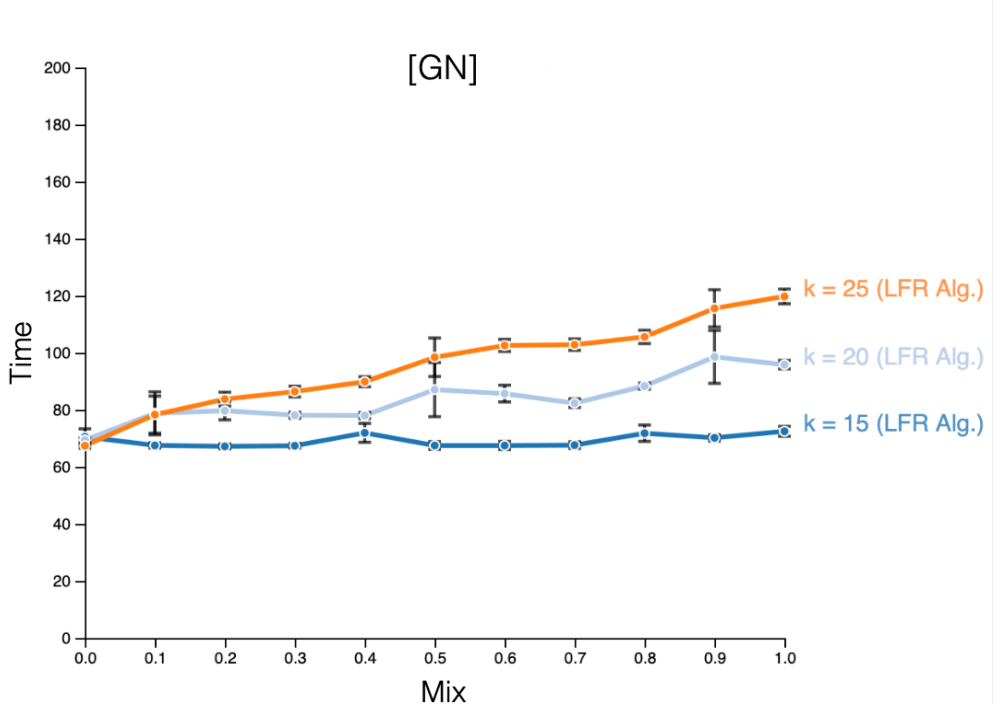
$$k_i - \gamma(v_i - k_i)$$

$L(M)$

Results

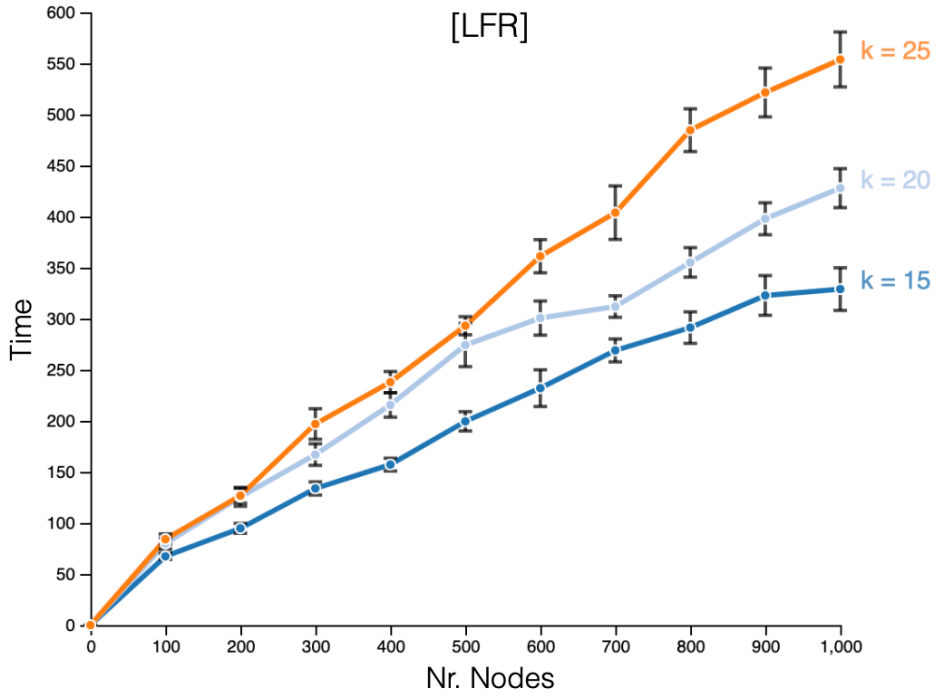
Benchmark Networks Algorithms

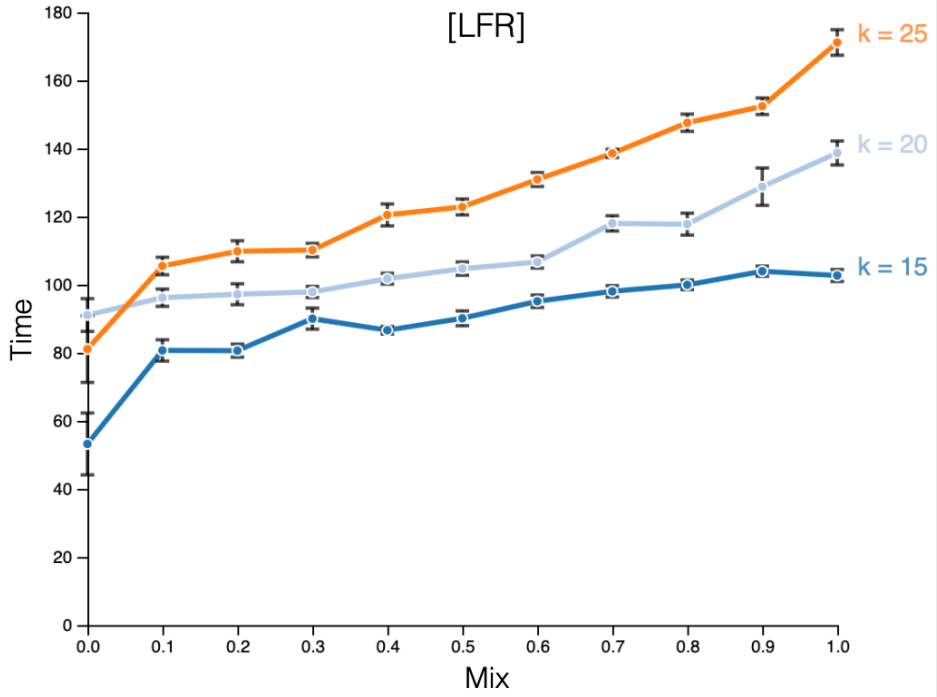




Results

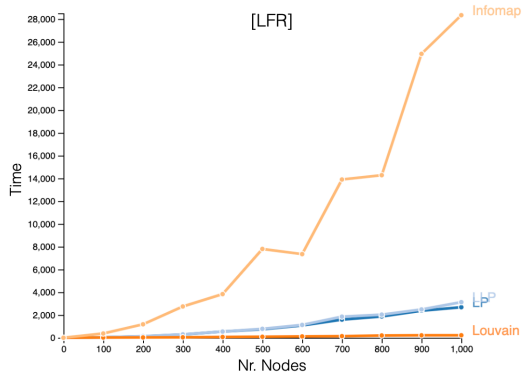
Benchmark Networks Algorithms



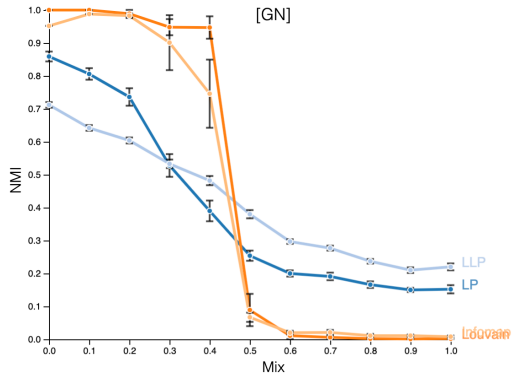


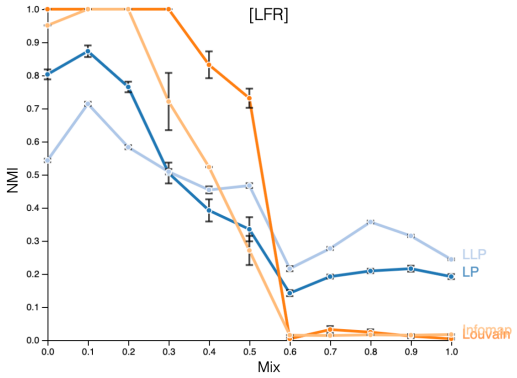
Results

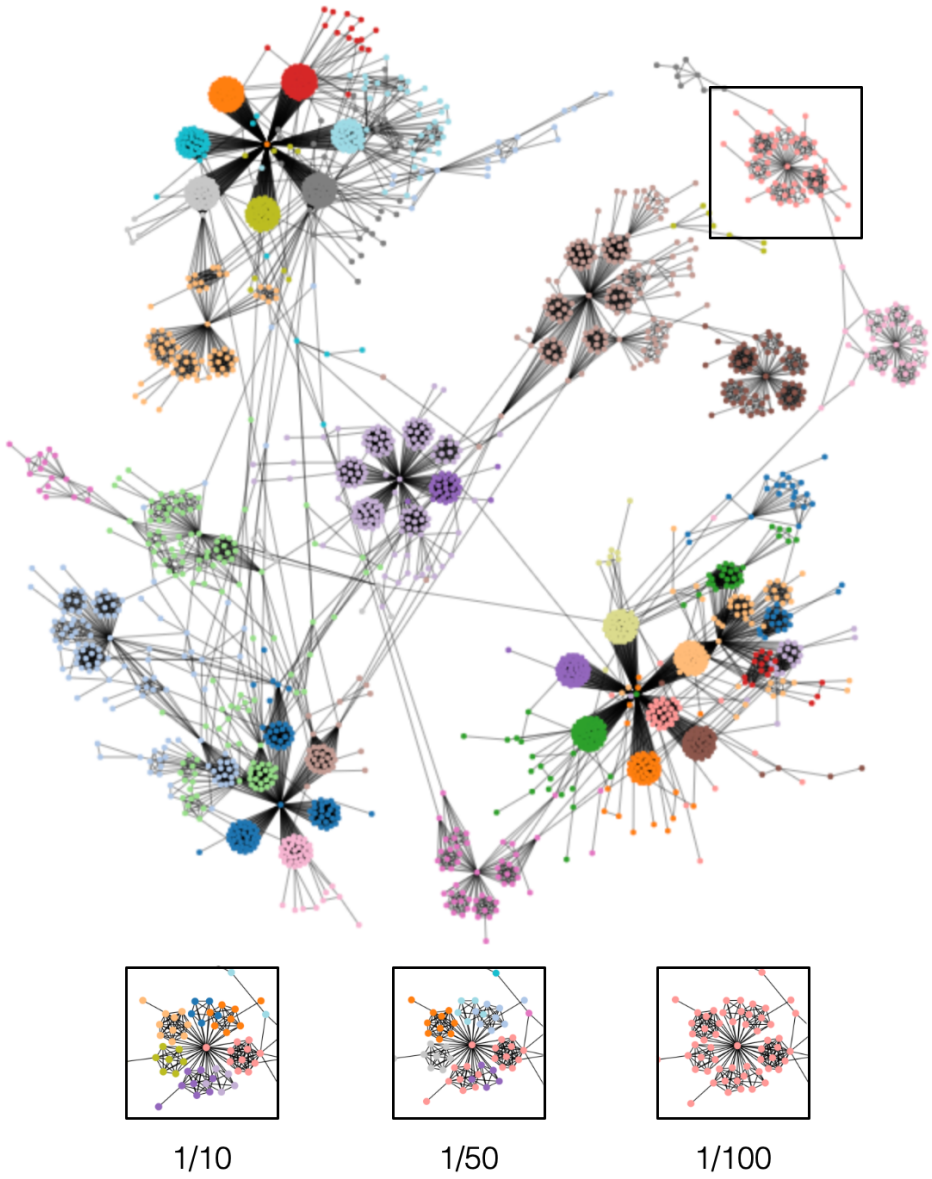
Staphylococcus aureus



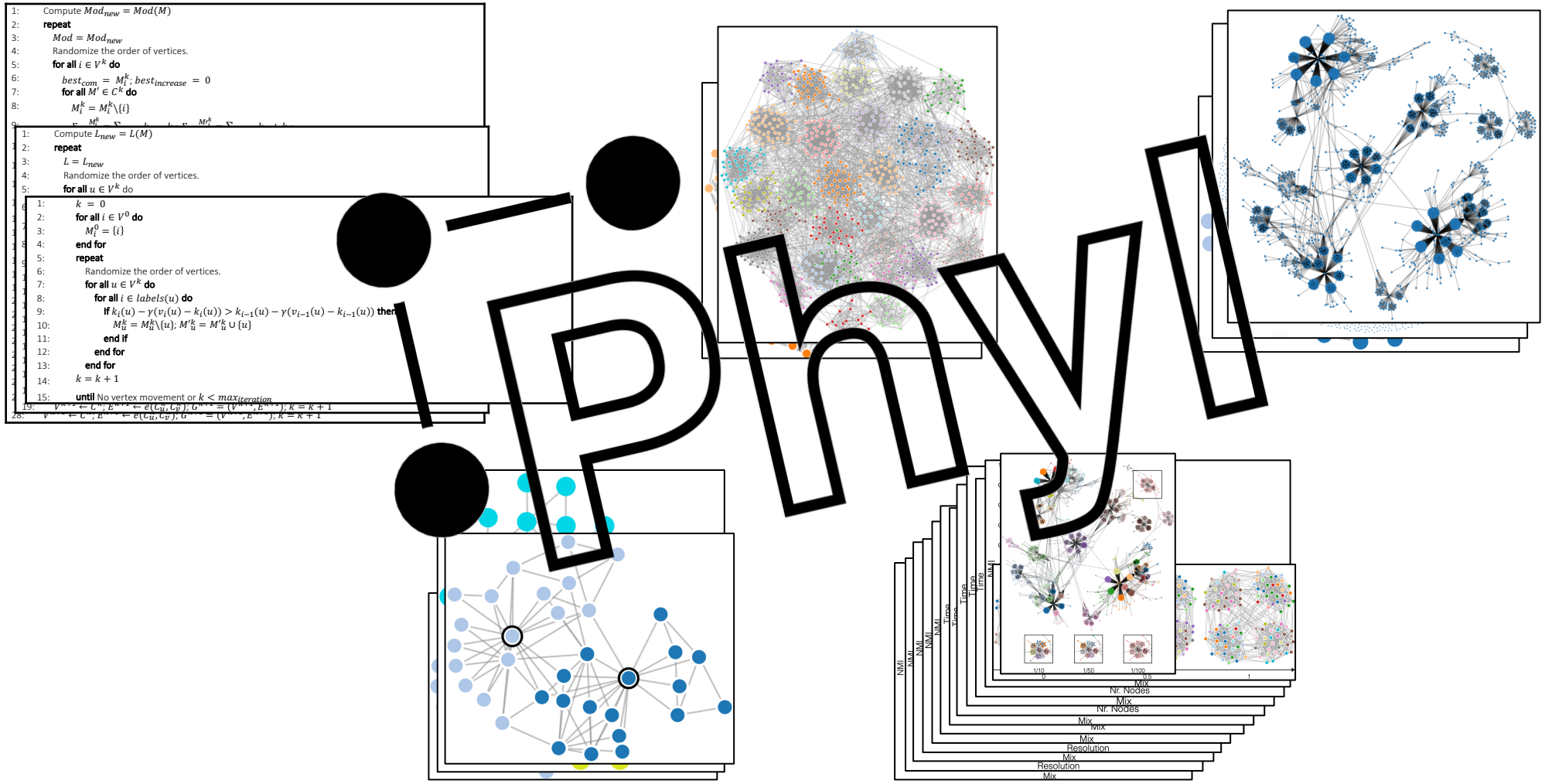
Louvain & D3.js (Canvas)



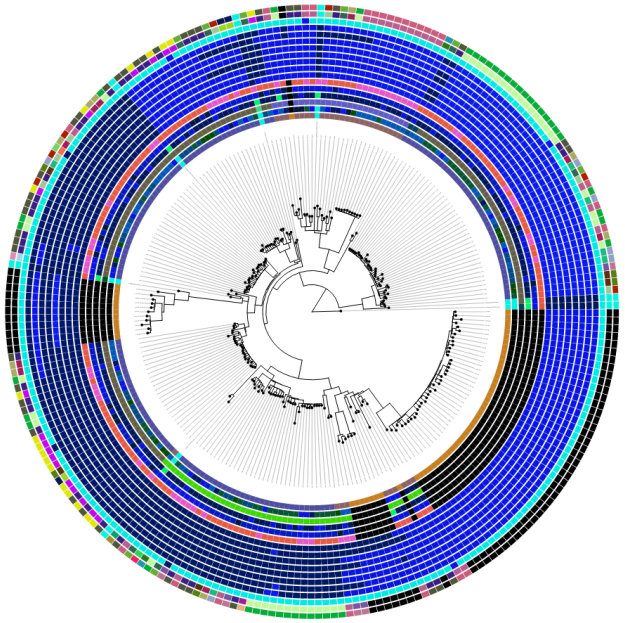


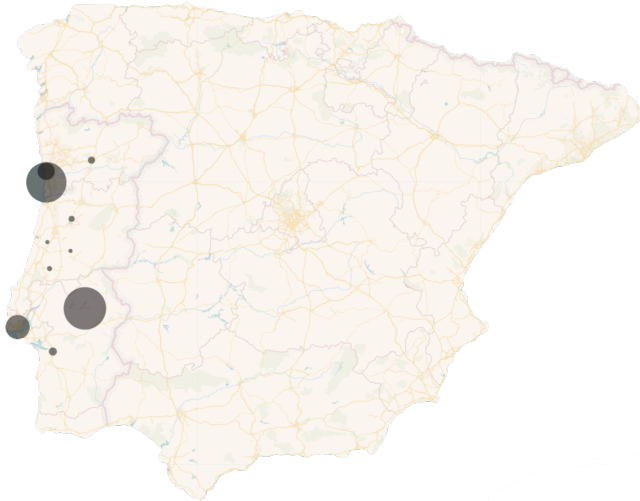


Conclusions

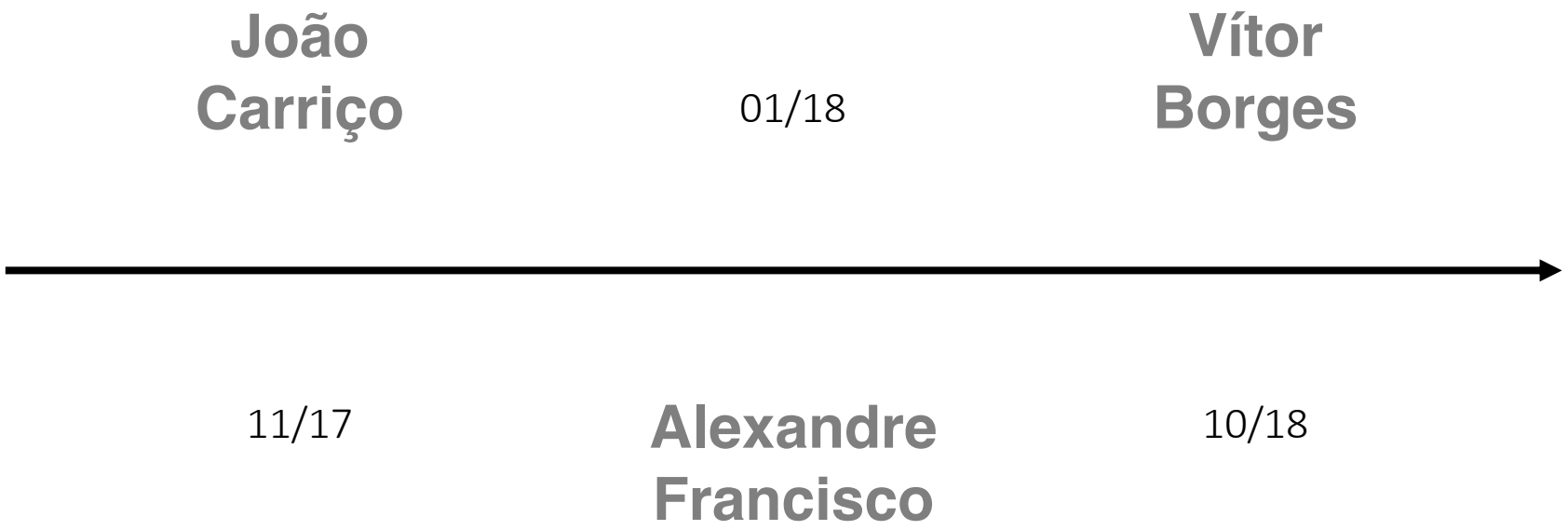


Future Work





Acknowledgments



References

[Slide 6] "The world's most valuable resource is no longer oil, but data," The Economist, 6 May 2017. [Online]. Available: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data. [Accessed 28 May 2019];

[Slide 6] "Ten threats to global health in 2019," World Health Organization, [Online]. Available: https://www.who.int/emergencies/ten-threats-to-global-health-in-2019. [Accessed 4 June 2019];

[Slide 6] "Top 10 Leading Causes of Death Globally," [Online]. Available: https://www.theatlas.com/charts/HkLaDreuW. [Accessed 12 May 2019];

[Slide 6] Y. Motro and J. Moran-Gilad, "Next-generation sequencing applications in clinical bacteriology," *Biomolecular Detection and Quantification,* vol. 14, p. 1–6, 2017;

[Slide 7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. (2008) P10008,* p. 12, 2008;

[Slide 7] M. Rosvall, D. Axelsson and C. T. Bergstrom, "The map equation," *The European Physical Journal Special Topics,* vol. 178, no. 1, pp. 13-23, 2009;

[Slide 7] P. Boldi, M. Rosa, M. Santini and S. Vigna, "Layered label propagation: a multiresolution coordinate-free ordering for compressing social networks," in *WWW '11 Proceedings of the 20th international conference on World wide web*, 2011;

[Slide 8] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 12, pp. 7821-7826, 2002;

[Slide 9] A. Lancichinetti, S. Fortunato and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical review. E, Statistical, nonlinear, and soft matter physics.,* vol. 78, no. 4, 2008;

[Slide 9] A. Lancichinetti, S. Fortunato and J. Kertesz, "Detecting the overlapping and hierarchical community structure of complex networks," *New Journal of Physics,* vol. 11, 2009;

[Slide 10] J. Yang and J. Leskovec, "Defining and Evaluating Network Communities based on Ground-truth," in *Proceedings of 2012 IEEE International Conference on Data Mining (ICDM)*, 2012;

[Slide 11] W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups," *Journal of anthropological research,* vol. 33, 1976;

[Slide 11] "Staphylococcus aureus MLST Databases," PubMLST, 5 June 2019. [Online]. Available: https://pubmlst.org/saureus/. [Accessed 5 June 2019];

[Slide 14] A. Lancichinetti and S. Fortunato, "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities," *Physical review. E, Statistical, nonlinear, and soft matter physics.,* vol. 80, 2009;