How tightly connected are communities?

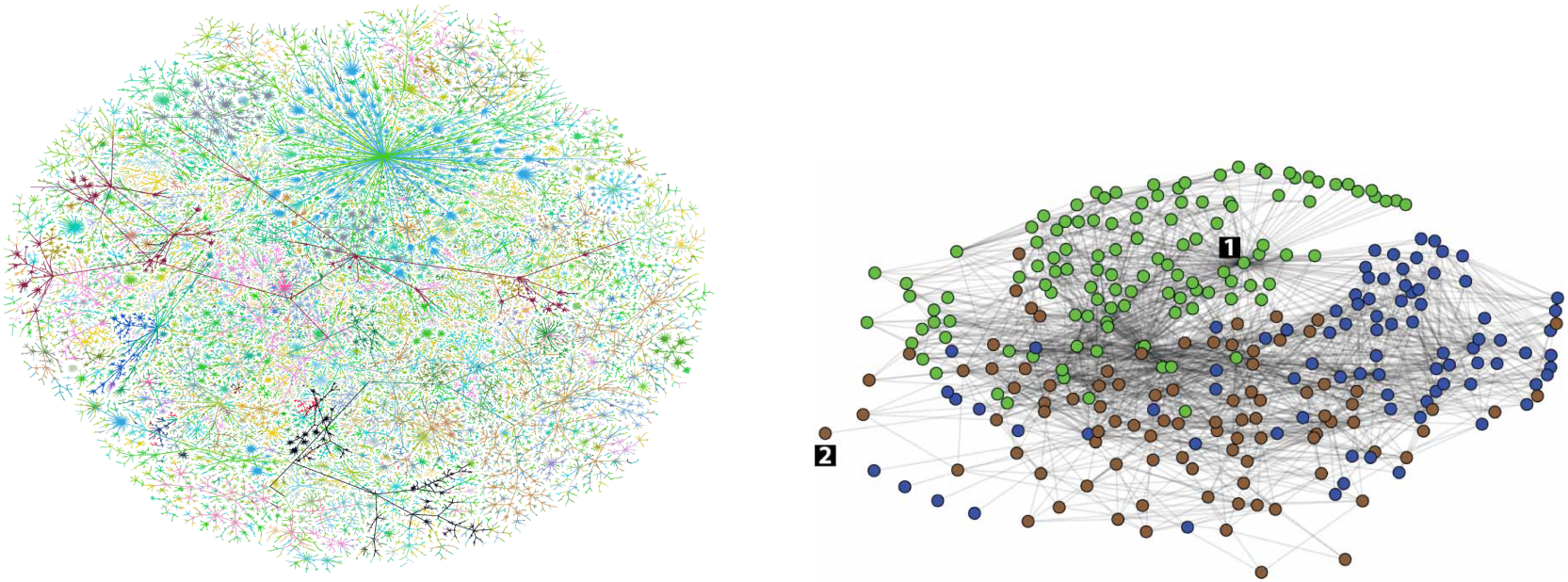
Minh Van Nguyen
minguyen@student.unimelb.edu.au
Supervisors: Michael Kirley and Rodolfo García-Flores

Department of Computing and Information Systems
The University of Melbourne

08 July 2013

Contents

Networks are everywhere



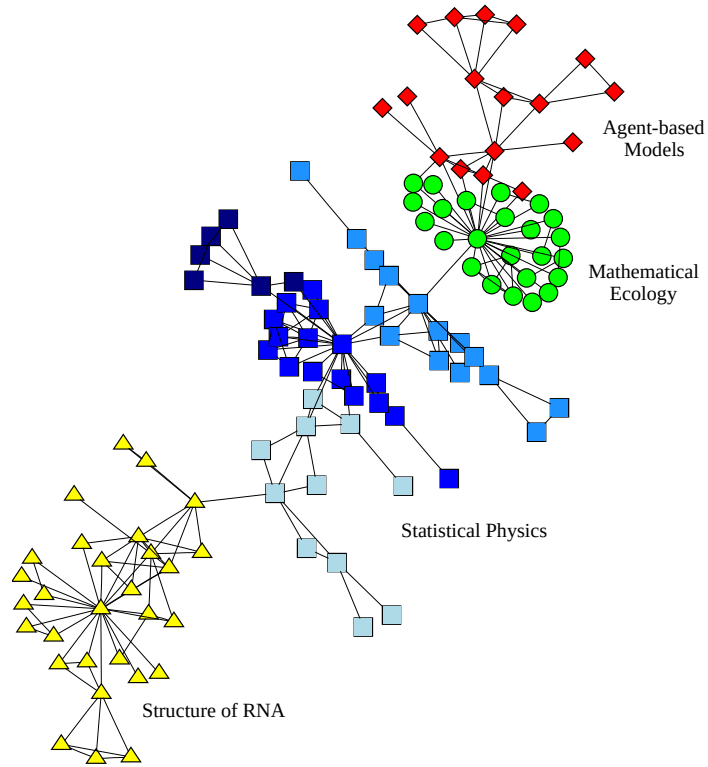
(a) internet at level of AS [6]

(b) neural network [1]

Figure 1: Technological and biological networks.

Motivations for studying communities include...

(a) market segmentation [4]

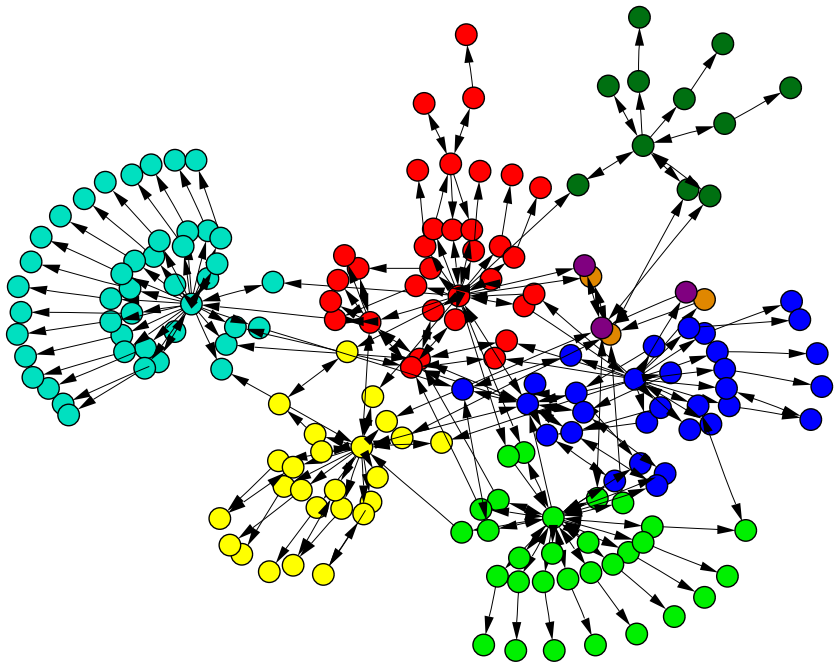


(b) social network analysis [3]

Figure 2: Communities in economic and social networks.

Motivations for studying communities include...

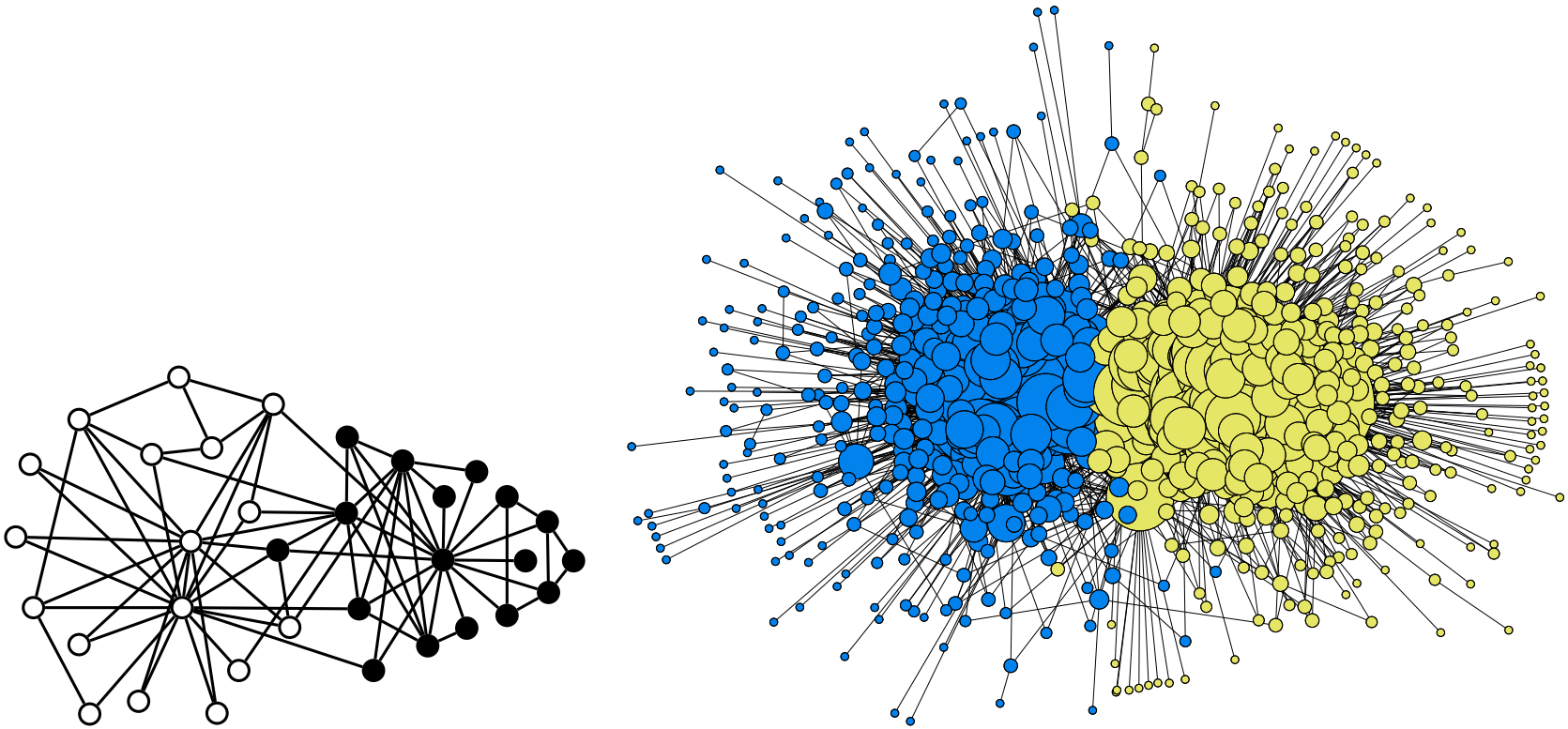
(a) organize computing clusters [2]



(b) analyze structure of WWW [7]

Figure 3: Communities in technological and information networks.

Communities are tightly inter-connected nodes



(a) karate club [8]

(b) political blogs [5]

Figure 4: Real-world networks with known community structures.

Hypotheses of community detection

1. Better connection between nodes *within* a community.
2. Few connection between nodes in *different* communities.

Research question

- How well inter-connected are nodes within a community?
- Define a compactness ratio that accounts for shortest paths between distinct pairs of nodes:

$$W^* = 1 - \frac{W(C)}{W(T)}$$

where

$$W(C) = \sum_{i<j} d_{ij}$$

and d_{ij} is the length of the shortest path between nodes i, j.

Results

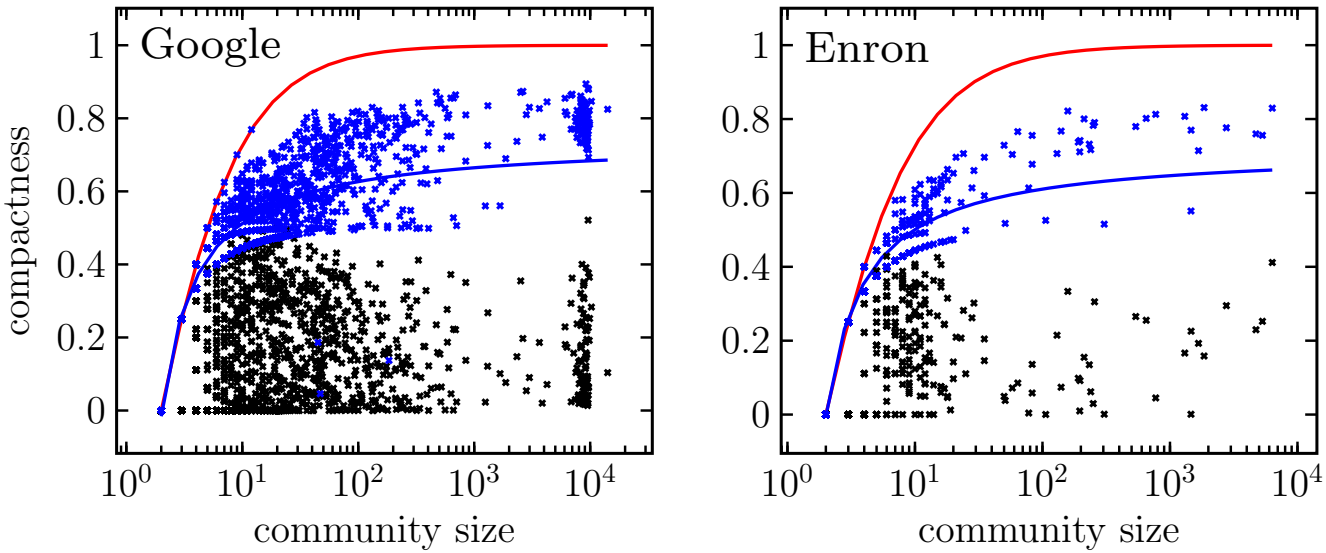


Figure 5: Compactness of communities in networks.

- black dot – compactness W^* of a community
- blue dot – ideal compactness $W^*_{K_n}$, i.e. when a community has all possible edges
- blue curve – model of the ideal compactness
- red curve – bound on both W^* and $W^*_{K_n}$

Shortcut links

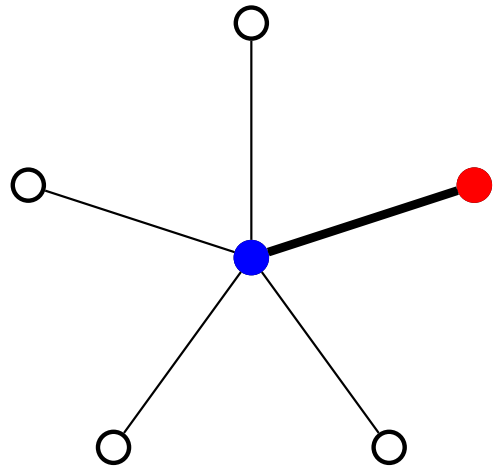


Figure 6: A node with many links (blue) connecting to a node with few links (red). The bold edge acts as a shortcut between the red node and other nodes.

Removing a shortcut edge is expected to reduce the compactness by a rate proportional to

$$-1/f_e^k$$

for some $k > 0$, where f_e is the fraction of edges removed.

Changes in compactness

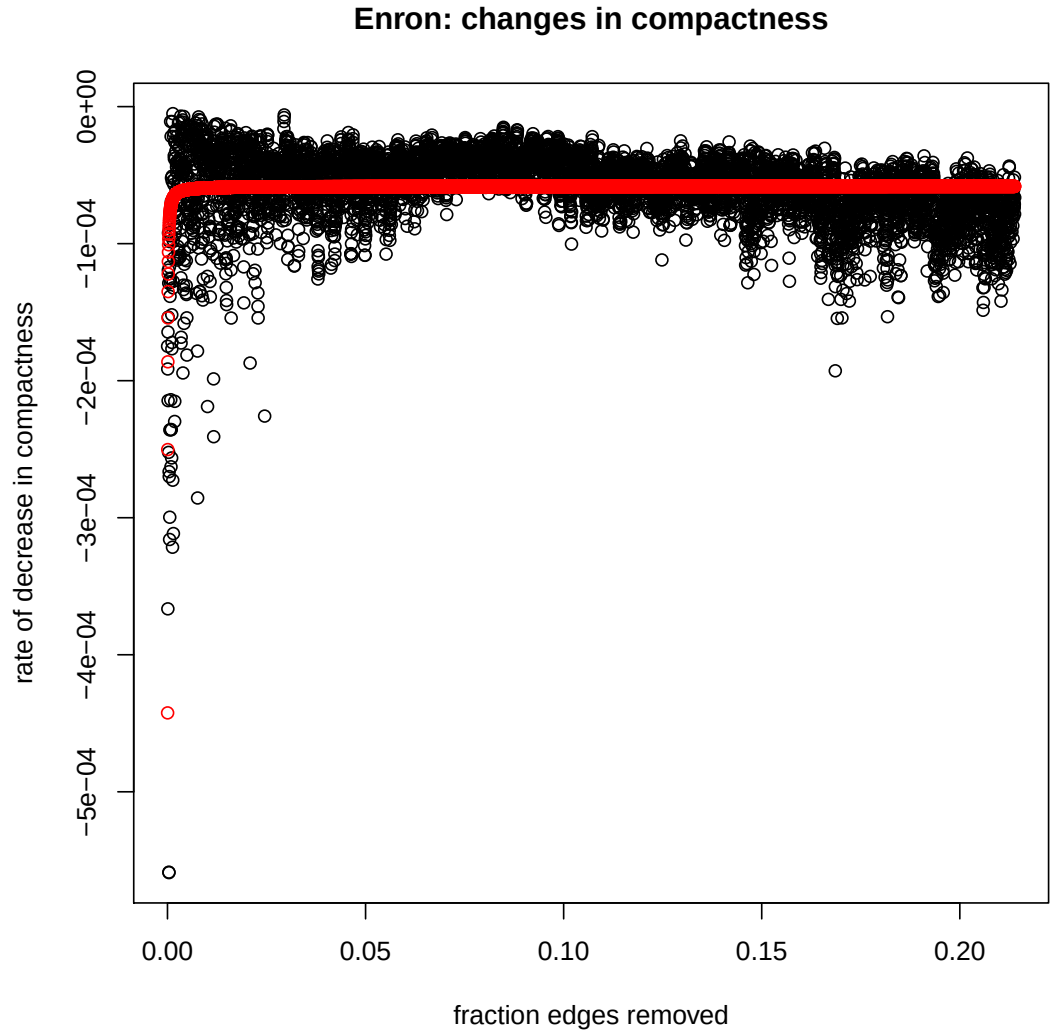


Figure 7: Largest community in the Enron email network. Change in compactness as a function of the fraction of edges removed.

Changes in compactness

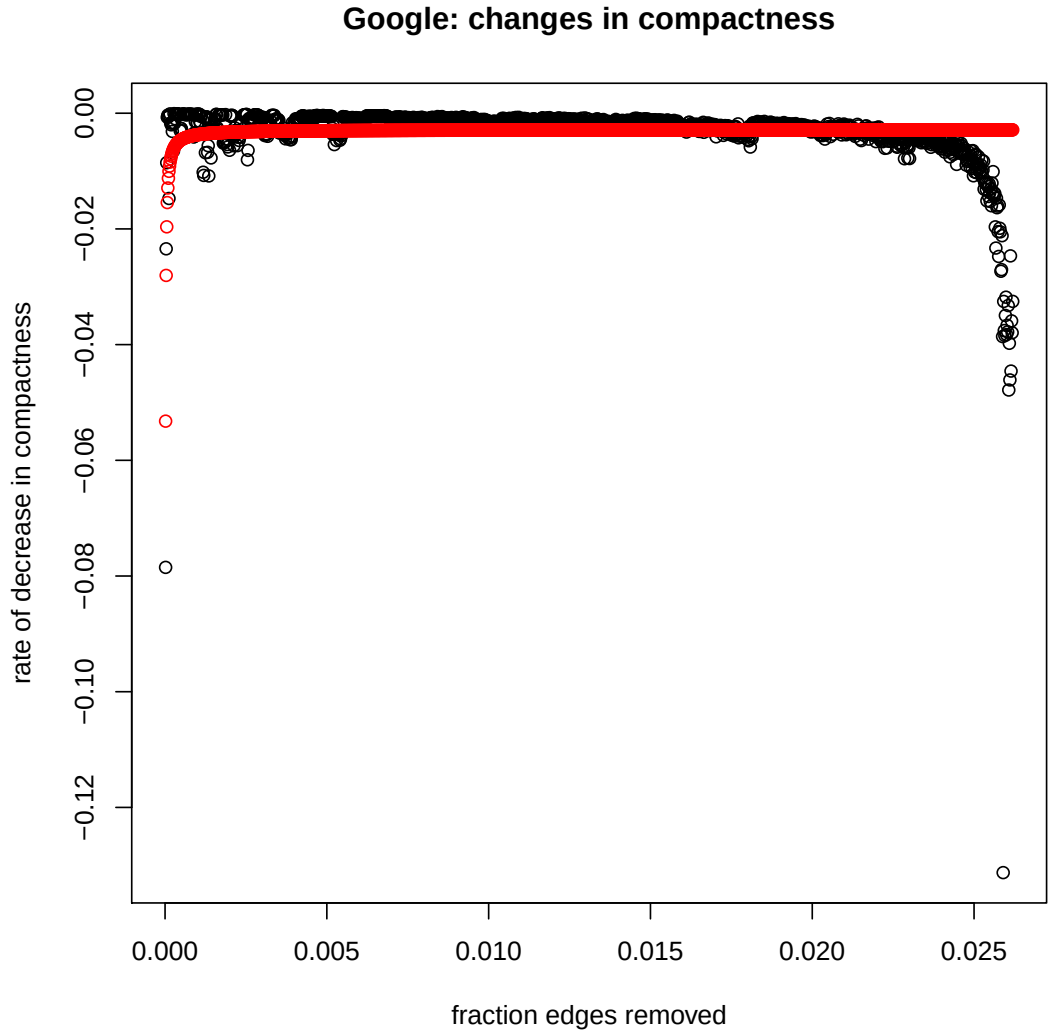


Figure 8: Largest community in the Google web graph. Change in compactness as a function of the fraction of edges removed.

Conclusion

- Address a fundamental hypothesis of community detection: Nodes within a community are better inter-connected among themselves.
- Compactness: Measure how well inter-connected are nodes in a community.
- Shortcut edges are important for increasing the inter-connectedness of nodes in a community.

Thank you

L. da Fontoura Costa, O. N. Oliveira Jr., G. Travieso, F. A. Rodrigues, P. R. V. Boas, L. Antiqueira, M. P. Viana, and L. E. C. Rocha.
Analyzing and modeling real-world phenomena with complex networks: a survey of applications.
Adv. Phys., 60:329–412, 2011.

U. de Valladolid.
Computational chemistry group, 11 August 2012.
http://www3.uva.es/ccg/facilities.html.

M. Girvan and M. E. J. Newman.
Community structure in social and biological networks.
PNAS, 99:7821–7826, 2002.

B. Isaacson.
Five pitfalls of market segmentation and how to avoid them, 11 May 2012.
http://www.mmrstrategy.com/think-tank/five-pitfalls-of-market-segmentation-and-how-to-avoid-them-2/.

B. Karrer and M. E. J. Newman.
Stochastic blockmodels and community structure in networks.
Phys. Rev. E, 83:016107, 2011.


M. E. J. Newman.
The structure and function of complex networks.
SIAM Rev., 45:167–256, 2003.


M. E. J. Newman and M. Girvan.
Finding and evaluating community structure in networks.
Phys. Rev. E, 69:026113, 2004.


W. W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthrop. Res., 33:452–473, 1977.