Community evolution in dynamic networks

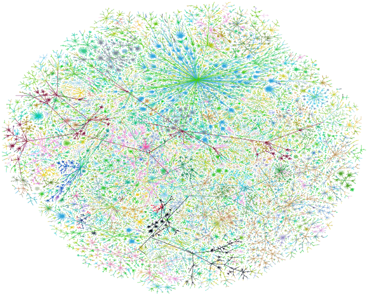
Minh Van Nguyen
minguyen@student.unimelb.edu.au
Supervisors: Michael Kirley and Rodolfo García-Flores

Department of Computing and Information Systems
The University of Melbourne

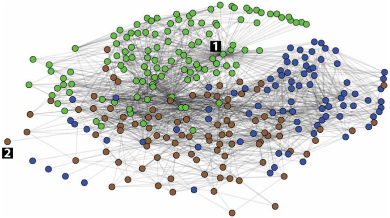
17 August 2012

Contents

Networks are everywhere



(a) internet at level of AS [8]

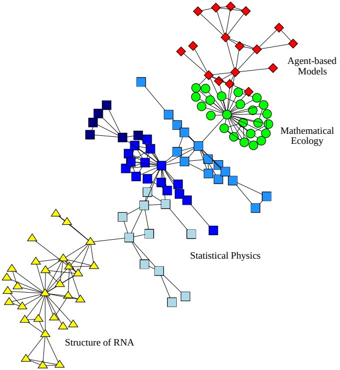


(b) neural network [2]

Figure 1: Technological and biological networks.

Motivations for studying communities include...

(a) market segmentation [6]

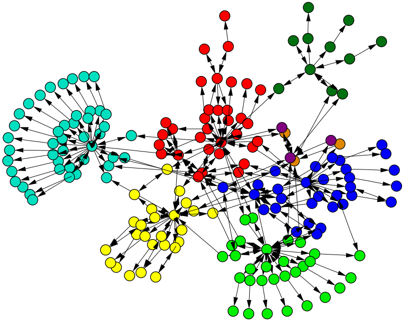


(b) social network analysis [5]

Figure 2: Communities in economic and social networks.

Motivations for studying communities include. . .

(a) organize computing clusters [3]



(b) analyze structure of WWW [10]

Figure 3: Communities in technological and information networks.

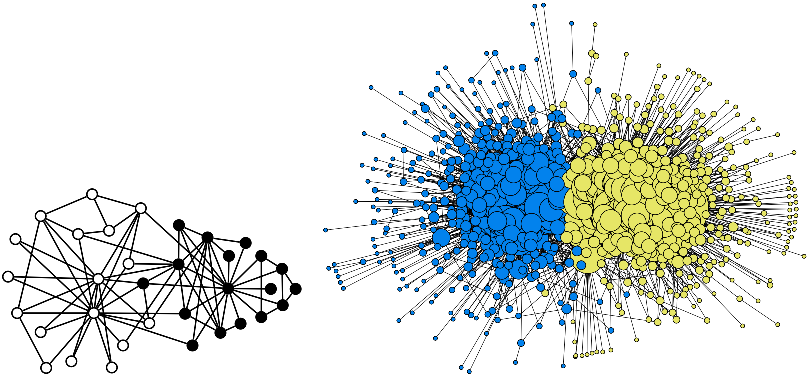
Research questions

1. What characteristics of community evolution are common across information, social, and technological networks?

2. How do we model and make predictions about communities in real-world networks?

Figure 4: Our research pipeline.

Communities are tightly connected nodes



(a) karate club [14] (b) political blogs [7]

Figure 5: More connections between nodes within community than to nodes outside of community.

Many definitions of communities

- Newman modularity Q [9] defined as

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

  The quantity m is the number of edges, A_{ij} is the weight of the edge (i, j), and k_i is the sum of weights of all edges attached to i. The delta function $\delta(c_i, c_j) = 1$ if i and j belong to the same community, and $\delta(c_i, c_j) = 0$ otherwise.
- q-state Potts model [13]
- clique percolation [12]
- and many more definitions [4]

Community tracking as a problem of object matching

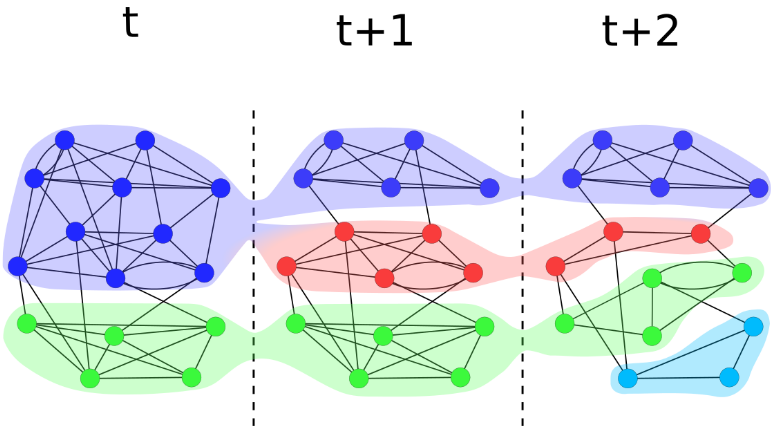
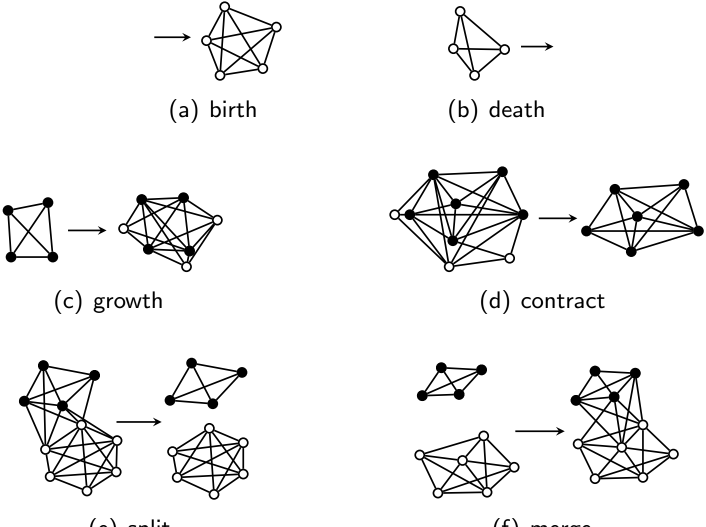


Figure 6: Infer communities across time [1]. We quantify continuity across time via the Jaccard coefficient $J(A, B) = |A \cap B| / |A \cup B|$.

Communities evolve according to a life-cycle



(a) birth (b) death

(c) growth (d) contract

(e) split (f) merge

Figure 7: Events in the life-cycle of communities [11].

Datasets on real-world dynamic networks

Scientific collaboration

- arXiv — physics, computer science, maths
- GP — genetic programming

Autonomous systems of the internet

- DIMES — similar to RouteViews
- Katrina — subset of RouteViews around Hurricane Katrina

Communities mostly have power-law lifespan ...

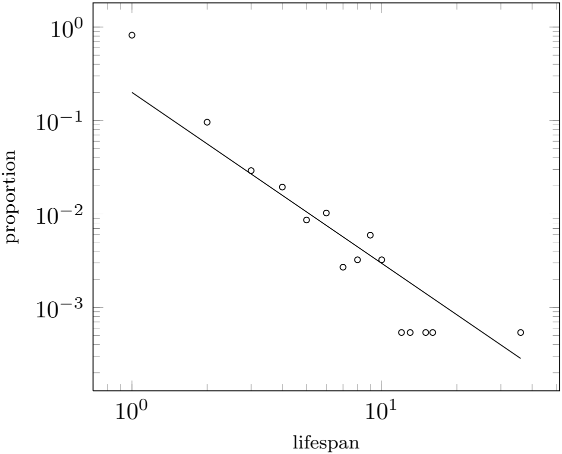


Figure 8: DIMES: lifespan follows a power-law of the form $\ell \sim k^{-\gamma}$.

. . . and sometimes exponential lifespan

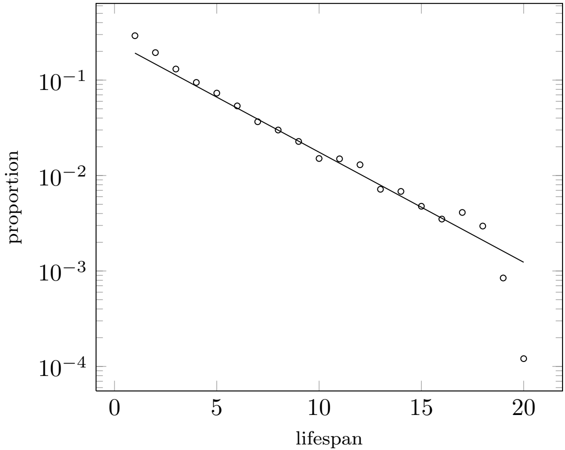
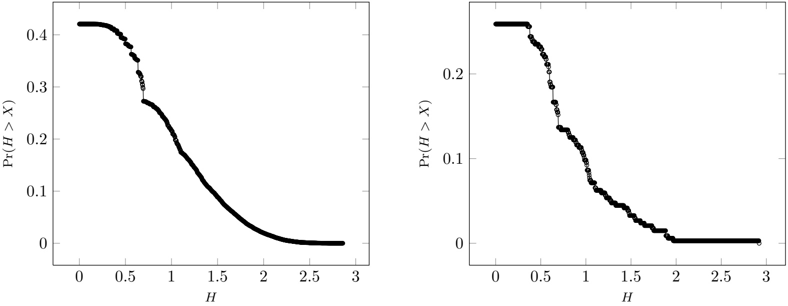


Figure 9: arXiv: lifespan follows an exponential law $\ell \sim \exp(-\lambda k)$.

Growth patterns can be highly unpredictable

$$\text{entropy of lifetime growth} = H = -\sum_k \Pr(k) \log \Pr(k)$$



(a) arXiv (b) DIMES

Figure 10: A stagnant majority and a tiny minority of "super growers".

Simulate a network with split and merge events

- node represents a community
- start with c communities at time $t = 1$
- add m communities at each time $t \geq 2$
- p is probability of split or merge at time t

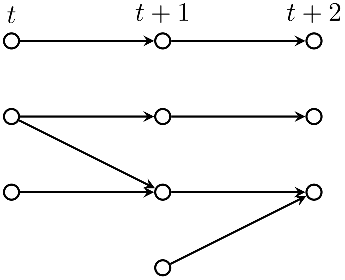


Figure 11: Communities in one time step split or merge with communities in the next time step.

Communities have small probability of split or merge

- window of interest is $p \in [0, 0.3]$

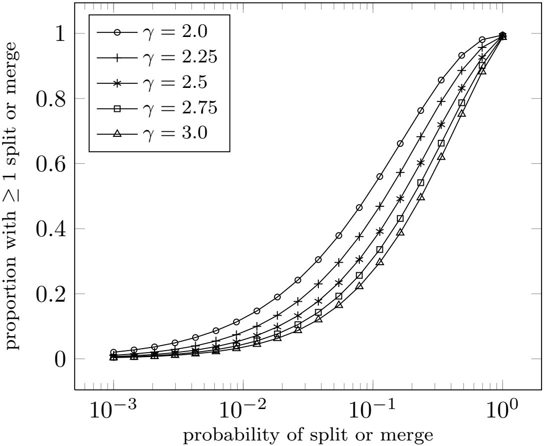


Figure 12: The proportion of communities with at least one split or merge as a function of *p*.

Future work

Figure 13: Our research pipeline.

Future work

Community detection

- Devise a fast algorithm to detect community without using Newman modularity.

Modelling & prediction

- What is an underlying process that is responsible for community growth and contraction?
- Devise a constructive model of community evolution.

Real-world datasets

- Information networks: citation networks, links in web pages
- Social networks: online groups, collaboration networks
- Technological networks: the internet, software networks

Thank you

T. Aynaud, J.-L. Guillaume, Q. Wang, and E. Fleury.
Communities in evolving networks: definitions, detections and analysis techniques, March 9 2011.
www-rp.lip6.fr/~magnien/DynGraph/Docs/art_communities.pdf.

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.

S. Fortunato.
Community detection in graphs.
Phys. Rep., 486:75–174, 2010.

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.
Analysis of weighted networks.
Phys. Rev. E, 70:056131, 2004.

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

G. Palla, A.-L. Barabási, and T. Vicsek.
Quantifying social group evolution.
Nature, 446:664–667, 2007.

G. Palla, I. Derényi, I. Farkas, and T. Vicsek.
Uncovering the overlapping community structure of complex networks in nature and society.
Nature, 435:814–818, 2005.

J. Reichardt and S. Bornholdt.
Statistical mechanics of community detection.
Phys. Rev. E, 74:016110, 2006.

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