Geography and community in complex networks

Vincent Blondel
LIDS, MIT
On leave from UCLouvain (Belgium)
Panel on Theoretical Challenges
Moderator: M. Vidyasagar
Panel: Peter Caines, Jan van Schuppen

3. System approximation in the 2-norm
A.C. Antoulas

16. A stabilization problem
Roger Brockett

25. Input-output gains of switched linear systems
J. P. Hespanha, A. S. Morse

40. Control-Lyapunov functions
Eduardo D. Sontag

46. Shift policies in QR-like algorithms and feedback control of self-similar flows
Paul Van Dooren, Rodolphe Sepulchre

50. Lyapunov theory for high order differential systems
Jan C. Willems

51. Performance lower bound for a sampled-data signal reconstruction
Yutaka Yamamoto, Shinji Hara

77. Important computational complexity problems in systems and control
Peter Caines
Panel on Theoretical Challenges
Moderator: M. Vidyasagar
Panel: Peter Caines, Jan van Schuppen

«Clearly, the most important and pressing problems in system and control are problems related to computational complexity. These are problems that have had an enormous impact in control and will continue to do so for the next decades.»

Peter Caines at SontagFest, May 2011

77. Important computational complexity problems in systems and control
Peter Caines
Mobile Data: A Gold Mine for Telcos
A snapshot of our activities, cell phone data attracts both academics and industry researchers.

By Tom Simonite
THURSDAY, MAY 27, 2010

Cell phone companies are finding that they're sitting on a gold mine—in the form of the call records of their subscribers.

Researchers in academia, and increasingly within the mobile industry, are working with large databases showing where and when calls and texts are made and received to reveal commuting habits, how far people travel for public events, and even significant social trends.

With potential applications ranging from city planning to marketing, such studies could also provide a new source of revenue for the cell phone companies. "Because cell phones have become so ubiquitous, mining the data they generate can really revolutionize the study of human behavior," says Ramón Cáceres, a lead researcher at AT&T's research labs in Florham Park, NJ.

If you were an AT&T subscriber and were near Los Angeles or New York between March 15 and May 15 last year, there's a 5 percent chance that your data was crunched by Cáceres and his colleagues in a study of the travel habits of the company's subscribers. The
Workshop on the Analysis of Mobile Phone Networks

A satellite workshop to NetSci 2010
Tuesday, May 11, 2010
MIT, Cambridge, MA

NetMob2011
Given the success of NetMob2010, we consider the possibility of organizing a NetMob2011. If you wish to be included on the NetMob mailing list, please send an email to sympa2@listes.uclouvain.be with "subscribe netmob yourname" in the subject line (where "yourname" is your first and last name). You can also subscribe/unsubscribe by going to https://listes-2.sipr.ucl.ac.be/sympainfo/netmob.

Introduction
Mobile phone datasets have become widely available in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, build trust, create markets and develop social interactions. Mobile phone data is also helping us understand complex processes such as the spread of information and viruses or transportation and the use of urban infrastructures.

This workshop will consist of a number of contributed talks on the analysis of mobile phone networks. The workshop format is flexible: no registration fees, a simplified submission procedure, and the possibility to present recent results or results submitted elsewhere.

Practical information
Date: Tuesday May 11, 2010 (this is the day prior to the conference NetSci).

Location: On the sixth floor of the newly built Media Lab (building E14 on MIT campus. map available here).

Registration: Attendance is free of charge but, due to limited seating, registration is compulsory. If you wish to register please send an email to netmob@uclouvain.be. Registration will be processed on a first-come first-serve basis. Although there is no registration fee for the workshop, participants are of course encouraged to also participate (and register) in the NetSci conference.

We have have received an unexpectedly large number of registrations to the workshop. The workshop has been moved to a larger space (the multi media hall of the Media Lab). All those who have registered by sending an email or through the NetSci website are welcome to attend.

Submissions
All contributions that deal with the analysis of mobile phone datasets are welcome.

Authors are invited to submit an abstract (one to three pages) by the deadline of March 5, 2010. Submissions should include the title, author(s), affiliation(s) and e-mail address(es) on the first page. There will be no published proceedings; the material submitted to the workshop may also be submitted elsewhere.

Electronic submission of manuscripts in PDF format is required. Please send your manuscript directly to netmob@uclouvain.be by March 5, 2010.

The evaluation of submitted abstracts will be organized by the scientific committee and decisions will be made by March 26, 2010. Once an abstract has been accepted for presentation, at least one author is required to attend the workshop and present the paper. In case too many abstracts are selected, some of these may be moved to a special session taking place the next day at the NetSci 2010 conference.

Program
The program is available here (PDF format).

Book of abstracts
The book of abstracts is available here (5.5 MB, PDF format).

Scientific committee
Chair: Vincent Blondel, UCLouvain (Belgium)
Laszlo Barabasi, Northeastern University
Bob Chiu, Apple, Inc.
Recent technological and mathematical developments have opened the possibility to considerably improve our understanding of how information flows and decisions are made in large social networks. This interdisciplinary workshop, we bring together researchers from different communities working on information propagation and decision making in social networks to investigate both rigorous models that highlight capabilities and limitations of such networks as well as empirical and simulations studies of how people exchange information, influence each other, make decisions and develop social interactions.

This workshop is being organized by the Laboratory for Information and Decision Systems.

**Organizers**

- Vincent Blondel, UCLouvain (Belgium) and LIDS, MIT
- Munther Dahleh, LIDS, MIT
- Asu Ozdaglar, LIDS, MIT
- John Tsitsiklis, LIDS, MIT

**Scientific committee**

- **Chair:** Vincent Blondel, UCLouvain (Belgium) and LIDS, MIT
- Daron Acemoglu, Economics, MIT
- Sinan Aral, New York University
- Albert-László Barabási, Northeastern University
Mobile Data: A Gold Mine for Telcos

A snapshot of our activities, cell phone data attracts both academics and industry researchers.

By Tom Simonite

THURSDAY, MAY 27, 2010

Cell phone companies are finding that they’re sitting on a gold mine—in the form of the call records of their subscribers.

Researchers in academia, and increasingly within the mobile industry, are working with large databases showing where and when calls and texts are made and received to reveal commuting habits, how far people travel for public events, and even significant social trends.

With potential applications ranging from city planning to marketing, such studies could also provide a new source of revenue for the cell phone companies. “Because cell phones have become so ubiquitous, mining the data that accounts can really revolutionize the way we do business,” says Ramón Cáceres, a researcher at Phone Labs in Florham Park, N.J.

This network shows the calls between around two million users in Belgium; each dot represents a connected group of people.
Large networks

- Web graph
- Internet graph
- Email exchange networks
- Blogs networks
- Citation networks
- Social networking sites (facebook, linkedIn, etc)
- Instant messages
- Mobile phone networks
- Twitter
- Collaboration graphs
- ...

Communities in networks
Why look for communities?

• Visualisation
• Structural organisation of the network
• Analysis (information propagation, robustness, cohesive,...)
• Time-evolution
Community detection in graphs
Santo Fortunato*

Complex Networks and Systems, Lagrange Laboratory, ID Foundation, Viale S. Severo 65, 10133, Turin, Italy

A R T I C L E   I N F O

Article history:
Received 5 November 2009
Accepted online 4 December 2009
Available online 5 December 2009

Keywords:
Graphs
Communities
Statistical physics

ABSTRACT

The modern science of networks has brought significant advances to our understanding of complex systems. One of the most relevant features of graphs representing real systems is community structure, or clustering, i.e., the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. Such clusters, or communities, can be considered as fairly independent compartments of a graph, playing a similar role like, e.g., the tissues or the organs in the human body. Detecting communities is of great importance in sociology, biology and computer science, disciplines where systems are often represented as graphs. This problem is very hard and not yet satisfactorily solved even though the enormous effort of a large interdisciplinary community of scientists working on it in the past few years. We attempt a thorough exposition of the topic, from the definition of the main elements of the problem, to the presentation of some of the most powerful methods developed. With a special focus on techniques designed by statistical physicists, from the discussion of crucial issues like the significance of clustering and how methods should be tested and compared, up to the description of applications to real networks.

Contents

1. Introduction ........................................ 76
2. Communities in real-world networks ......... 78
3. Elements of community detection .......... 82
   3.1. Computational complexity ................... 83
   3.2. Communities .................................... 83
       3.2.1. Basics ..................................... 83
       3.2.2. Local definitions .......................... 84
       3.2.3. Global definitions ......................... 85
       3.2.4. Definitions based on vertex similarity .. 86
   3.3. Partitions ...................................... 87
       3.3.1. Basics ..................................... 87
       3.3.2. Quality functions: Modularity ........... 88
   4. Traditional methods ............................ 90
       4.1. Graph partitioning ........................... 90
       4.2. Hierarchical clustering ...................... 93
       4.3. Partitional clustering ......................... 93
       4.4. Spectral clustering ........................... 94
   5. Divisive algorithms ............................ 96
       5.1. The algorithm of Girvan and Newman .... 97
       5.2. Other methods ................................ 99

* Tel.: +39 011 6003900; fax: +39 011 6000409.
E-mail address: fortunato@ph.la.infn.it.
0370-1573/$ – see front matter © 2009 Elsevier B.V. All rights reserved.

INTRODUCTION: NETWORKS AND COMMUNITIES

"But although it is a matter of history; mathematical machinery only ever in so far as its investigation in thermodynamics, is same essentially worthy of Thermodynamics, development, both on account of the elegance and simplicity of its principles, and because it yields new results and places old results in a new light, it is a mathematical abstraction apart from the world of thermodynamics."


For an abstract perspective, the world network is used as a synonym for a mathematical graph. However, in some scientific fields, a network is defined as much as an ensemble (i.e., 12, 25, 44, 83, 108, 135, 136). A network is a collection of nodes, or vertices of a network, and a pair of nodes can be connected by a link (or edge) that signifies some social interaction or tie between them (see Figure 1). Notice that the term network is used in different scientific disciplines, and in a variety of contexts, to describe the dynamics of interacting and interacting networks. For instance, in biology, the dynamics of networks are obtained through averaging over smaller networks of interest, with some set of microscopic interactions. However, in many situations, links can also be assigned to the nodes of a network, with some set of microscopic interactions. In this case, the term network can be misleading. The presence of modular structure may also mislead in the case of random networks. In such networks, the term community means that the term community may be sparse. When there is large variation among the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law. In the case of random networks, node degrees are distributed according to the normal distribution, but in many man-made and biological networks, the degree distribution follows a power-law.
Geography and community in complex networks

Louvain method for community detection (and modularity)
[VB, Guillaume, Lambiotte, Lefèvre, 2008]

Communities in a mobile phone network
[Lambiotte, VB et al., 2009]

Geography in community detection
[Krings, Calabrese, Ratti, VB, 2009]
[VB, Krings, Thomas, 2010]
[Expert, Evans, VB, Lambiotte, PNAS, 2011]

Eigenvectors and communities
[Cucuringu, Vandooren, VB, 2011]
**Web: 118M/1G**

Divide and Conquer: Partitioning Online Social Networks
Josep M. Pujol, Vijay Erramilli, Pablo Rodriguez
arXiv, 2010
**Twitter: 2.4M/38M**

Mapping search relevance to social networks
Jonathan Haynes, Igor Perisic
Proceedings of the 3rd Workshop on Social Network Mining and Analysis, 2010
**Linkedin, 21M**

Tracking the Evolution of Communities in Dynamic Social Networks
Greene, D.; Doyle, D.; Cunningham, P.;
International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2010
**Mobile phone, 4M/100M**

Real World Routing Using Virtual World Information
Pan Hui, Sastry N.
International Conference on Computational Science and Engineering, 2009
**Flickr 1.8M/22M, LiveJournal 5.3M/77M, YouTube 1.1M/4.5M**

Community structure in audio clip sharing
Gerard Roma, Perfecto Herrera
International Conference on Intelligent Networking and Collaborative Systems, INCoS 2010
**Freesound**

Subject clustering analysis based on ISI category classification
Lin Zhang, Xinhai Liu, Frizo Janssens, Liming Liang and Wolfgang Glänzel
Journal of Informetrics, Volume 4, Issue 2, April 2010
**ISI 6M papers**
NetworkX

High productivity software for complex networks

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Quick Example

```python
>>> import networkx as nx
>>> G=nx.Graph()
>>> G.add_node("span")
>>> G.add_edge(1,2)
>>> print(G.nodes())
[1, 2, 'span']
>>> print(G.edges())
[(1, 2)]
```
Modularity

Quality of a partition of a network in communities

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

$$\text{Modularity} = \frac{\text{# edges in communities} - \text{expected # edges in communities}}{\text{total number of edges}}$$

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

\[
\frac{\text{# edges in communities} - \text{expected # edges in communities}}{\text{total number of edges}}
\]

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

\[
\frac{\text{# edges in communities} - \text{expected # edges in communities}}{\text{total number of edges}}
\]

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

\[
\frac{\text{# edges in communities} - \text{expected # edges in communities}}{\text{total number of edges}}
\]

\( m \) edges, \( d_i \) degree of node \( i \)

expected # edges between node \( i \) and node \( j \) = \( (d_i \cdot d_j)/(2m) \)

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

$\text{# edges in communities - expected # edges in communities} \over \text{total number of edges}$

$m = 25$, expect 0.3 edges

$m$ edges, $d_i$ degree of node $i$

expected # edges between node $i$ and node $j = (d_i d_j)/(2m)$

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

Expected $11.56$  $2.56$

$m$ nodes, $d_i$ degree of node $i$

expected # edges between node $i$ and node $j = (d_i d_j)/(2m)$

expected # edges in community $\sum_{i, j \in C} (d_i d_j)/(2m)$

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

\[
\frac{\text{# edges in communities} - \text{expected # edges in communities}}{\text{total number of edges}}
\]

Expected 11.56  2.56
Observed 15  6

\(m\) nodes, \(d_i\) degree of node \(i\)

expected # edges between node \(i\) and node \(j\) = \((d_i d_j)/(2m)\)

expected # edges in community \(Sum_{i,j \in C} (d_i d_j)/(2m)\)

[Newman, 2004]
Modularity

Quality of a partition of a network in communities

\[
\frac{\text{# edges in communities} - \text{expected # edges in communities}}{\text{total number of edges}} = \frac{(15 + 6) - (11.56 + 2.56)}{25} = 0.275
\]

Expected 11.56 2.56
Observed 15 6

\(m\) nodes, \(d_i\) degree of node \(i\)

expected # edges between node \(i\) and node \(j\) = \((d_i \cdot d_j)/(2m)\)

expected # edges in community \(\sum_{i,j \in C} (d_i \cdot d_j)/(2m)\)

[Newman, 2004]
Optimizing modularity is NP-complete
[Brandes, 2008]
[Caines, 2012]
The (greedy) Louvain method

Initially every node forms a community.

For every node $i$, insert node $i$ in a neighboring community that maximizes the resulting modularity gain.

Repeat until a local maximum is attained.

Construct the resulting network of communities and repeat the construction on the network of communities.

This way we construct a hierarchy of communities.
The (greedy) Louvain method

Hierarchy of communities

Simple to implement

Low computational complexity (not $n^2$ but $n \log n$).
118M nodes/1B links in 152mn
• Belgian phone call network
  – 6 months of communications
  – One Belgian main operator

• Network:
  – 2.6 M customers
  – 800M voice/text messages
  – language information (Dutch, English, French, German)
  – location information (ZIP)

[Lambiotte, VB et al., 2008]
Mobile phone network
Distribution of calls received

Number of customers

Number of calls received
Duration with distance
Connection with distance

FIG. 2: We plot the probability $P_d$ that two people living at a distance $d$ are connected by a link in a log-log scale. The dashed line is the power-law $d^{-2}$.

[Onnela, Arbesman, Barabasi, Christakis, 2010]

[Lambiotte, VB et al., 2008]
Communities in Belgium
"Oui, il faut se préparer à la fin de la Belgique"

P
dès de trois mois après les élections, la Belgique se retrouve de nouveau dans la crise après la démission, entérinée ce week-end, d’Elio di Rupo, chef de file du Parti socialiste francophone qui a renoncé à son tour à tenter de former un gouvernement. Le roi Albert II a accepté samedi soir la démission du socialiste wallon, qui n’a pu combler le fossé entre néerlandophones et francophones paralysant la vie politique belge depuis plus de trois ans.

Conséquence de ce blocage, le tabou de la scission de la Belgique commence à tomber dans le monde politique francophone, dont plusieurs représentants de premier plan ont ouvertement évoqué cette éventualité, dimanche 5 septembre, en raison des difficultés à s’entendre sur l’avenir du pays avec les Flamands.
Poelvoorde : « Gardons nos barbes jusqu'à ce que la Belgique se relève »

Rédaction en ligne
jeudi 13 janvier 2011, 00:09

Benoît Poelvoorde suggère à la gent masculine de ne plus se raser « jusqu'à ce que la Belgique se relève » et soit dotée d'un gouvernement, alors que le royaume traverse la plus longue crise politique de son histoire.
Geography and community in complex networks

Louvain method for community detection (and modularity)
[VB, Guillaume, Lambiotte, Lefèvre, 2008]

Communities in a mobile phone network
[Lambiotte, VB et al., 2009]

Geography in community detection
[Krings, Calabrese, Ratti, VB, 2009]
[VB, Krings, Thomas, 2010]
[Expert, Evans, VB, Lambiotte, PNAS, 2011]
Gravitational law of interactions

[Krings, Calabrese, Ratti, VB, 2008]
Géographie / Une étude redessine la Belgique en couleurs

Les flux GSM élargissent Bruxelles

L'essentiel

- Trois chercheurs de l'UCL ont analysé plus de 200 millions de communications.
- Bruxelles et cinq communes à l'extérieur sont francophones.

En tête, la Belgique administrative et régionale correspond à l'actuel réseau de télécommunication mobile. La Belgique administrative, régionale et géopolitique a été comparée à cinq parmi les six communes à l'extérieur. Les études de communication mobile et de téléphonie mobile ont été réalisées pour connaître les situations possibles de l'UCL.

30/09/10 21:54 - LE SOIR du 01/10/10 - p. 5
The Really Smart Phone

Researchers are harvesting a wealth of intimate detail from our cellphone data, uncovering the hidden patterns of our social lives, travels, risk of disease—even our political views.

By ROBERT LEE HOTZ

Apple and Google may be intensifying concerns by tracking where and when people use their mobile phones—but the true focus of consumer surveillance is taking shape as the cellphones at a weather-stained apartment complex in Cambridge, Mass.

For almost two years, Alex Pentland at the Massachusetts Institute of Technology tracked 60 families living in campus quads via sensors and software on their smartphones—recording their movements, relations, moods, health, calling habits and spending this wealth of intimate detail, he finds patterns of human behavior that could explain how millions of people interact at home and play.

Through these and other cellphone research projects, scientists are able to pinpoint "influencers," the people most likely to change others' minds. The data car...