introduction to *network analysis* (*ina*)

Lovro Šubelj

University of Ljubljana
Faculty of Computer and Information Science
spring 2019/20

announcements *5th week*

- *homework #1* due *next week*
- *homework #2* out in *two weeks*

- *project & report* details *today*
- *grand graph* challenge *review*

- *no lectures & labs this week*
- *feedback box* from *last week*

link *bridging*

introduction to *network analysis* (*ina*)

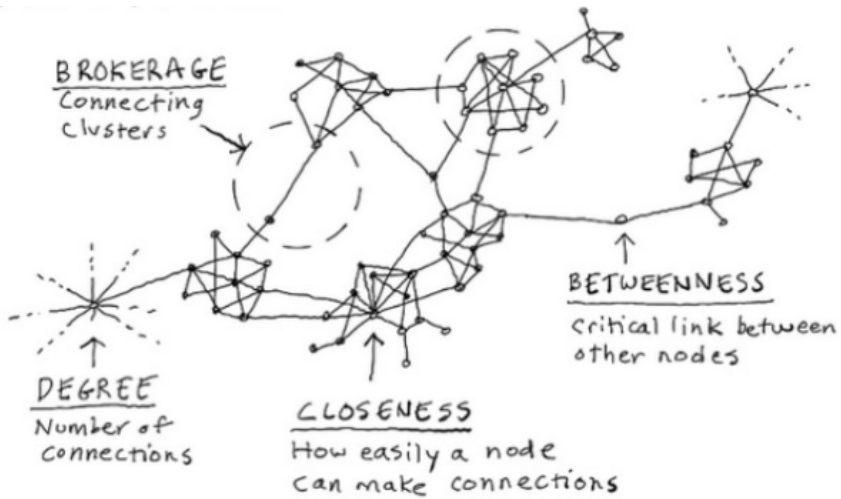
Lovro Šubelj
University of Ljubljana
spring 2019/20

bridging *measures*

which *links* are most *important*?

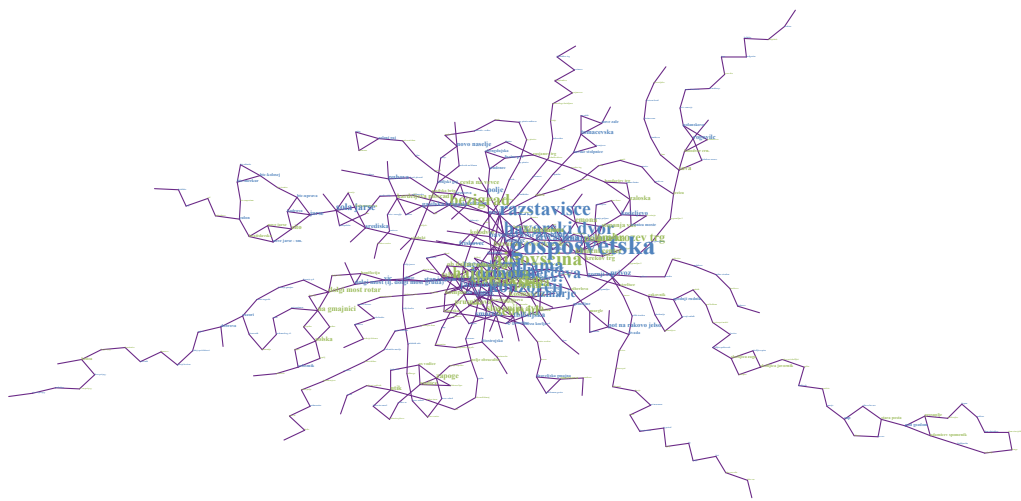
— *link bridging measures* for (*un*)*directed* networks
  – *betweenness-based* centrality [Fre77, FBW91, New05]

— *link embeddedness measures* for (*un*)*directed* networks
  – *topological overlap* measures [RSM+02, OSH+07, dNMB11]



networkology *LPP*

- partial *LPP public bus transport network*[*]
- $n = 416$ bus stops with $\langle k \rangle = 2.72$ connections
- *giant component* 95.4% nodes (6 components)
- *"small-world"* with $\langle C \rangle = 0.09$ and $\langle d \rangle = 14.26$
- *"scale-free"* with $\gamma = 2.43$ for cutoff $k_{min} = 2$



[*] reduced to largest connected component of simple undirected graph

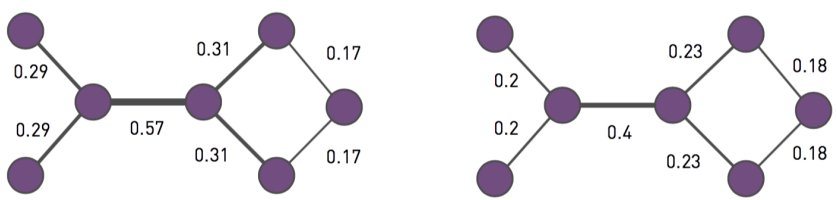
bridging *betweenness*

important *links* are *between other nodes*

— for (*un*)*directed* G *link betweenness* σ [Fre77] of $\{i, j\}$ is
  - g_{st} is number of *geodesic paths between* s and t
  - g_{st}^{ij} is number of *such geodesic paths through* $\{i, j\}$

$$\sigma_{ij} = \sum_{st \notin \{i,j\}} \frac{g_{st}^{ij}}{g_{st}}$$

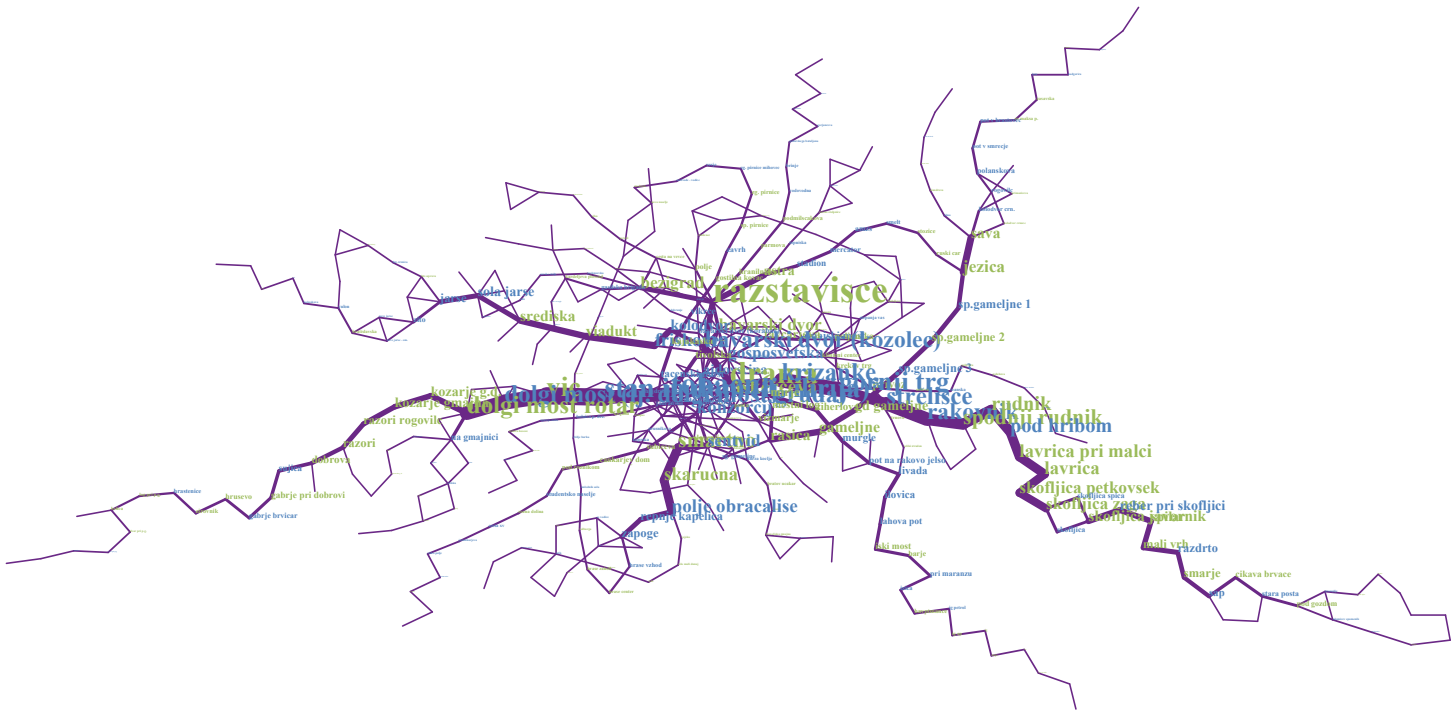
— σ considers *only geodesic paths* [FBW91, New05]



networkology *betweenness*

— *link betweenness* σ in partial LPP network[†]

— *highest* $\sigma_{ij} = 0.176n^2$ link is {*Vič, Stan in dom*}



[†] reduced to largest connected component of simple undirected graph

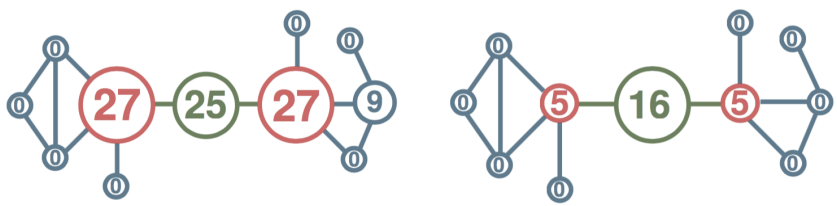
bridging *bridgeness*

important *links* are *bridges between nodes*

— for (*un*)*directed* G *link bridgeness* $\widetilde{\sigma}$ [JMK+16] of $\{i, j\}$ is
  - g_{st} is number of *geodesic paths between* s and t
  - g_{st}^{ij} is number of *such geodesic paths through* $\{i, j\}$

$$\widetilde{\sigma}_{ij} = \sigma_{ij} - \sum_{st \in \Gamma_i \cup \Gamma_j} \frac{g_{st}^{ij}}{g_{st}} = \sum_{st \notin \Gamma_i \cup \Gamma_j} \frac{g_{st}^{ij}}{g_{st}}$$

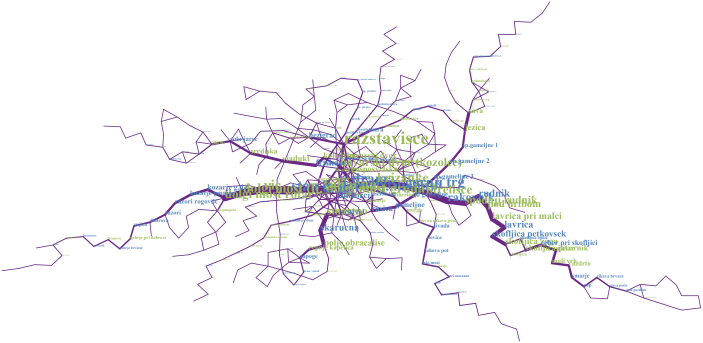
— σ mixes *local centers* with *global bridges* [JMK+16]



networkology *bridgeness*

— *link bridgeness* $\widetilde{\sigma}$ in partial LPP network[‡]

— *highest* $\widetilde{\sigma}_{ij} = 0.169n^2$ link is $\{$*Vič, Stan in dom*$\}$



‡ reduced to largest connected component of simple undirected graph

bridging *embeddedness*

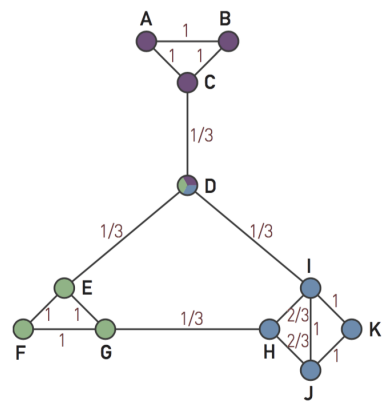
important *links* are *embedded between nodes*

— for *undirected* G *link embeddedness*[§] θ [OSH+07] of $\{i,j\}$ is
  – Γ_i is set of *neighbors* or *neighborhood* of i

$$\theta_{ij} = \frac{|\Gamma_i \cap \Gamma_j|}{k_i - 1 + k_j - 1 - |\Gamma_i \cap \Gamma_j|} \qquad \theta_{ij} = 0 \text{ for } k_i = k_j = 1$$

— μ*-corrected link embeddedness* $\widetilde{\theta}$ [dNMB11] of $\{i,j\}$ is
  – μ is *maximum* number of *triangles* over *links*

$$\widetilde{\theta}_{ij} = \frac{|\Gamma_i \cap \Gamma_j|}{\mu + \max(k_i,k_j) - 1 - |\Gamma_i \cap \Gamma_j|}$$

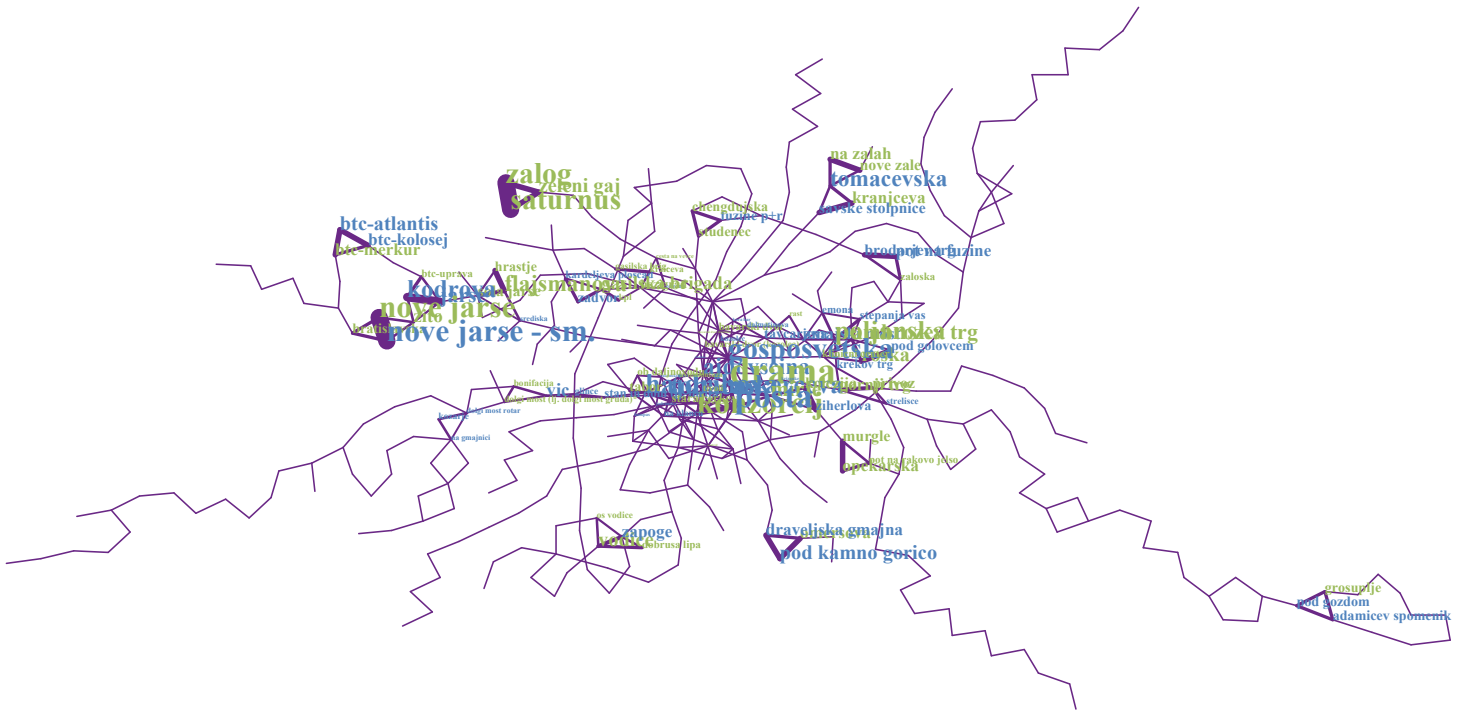


[§] θ & $\widetilde{\theta}$ better known as topological overlap indices/weights

networkology *embeddedness*

— *link embeddedness* θ in partial LPP network[¶]

— *highest* $\theta_{ij} = 1.0$ links are $\{$*Zalog, Saturnus*$\}$ *etc.*

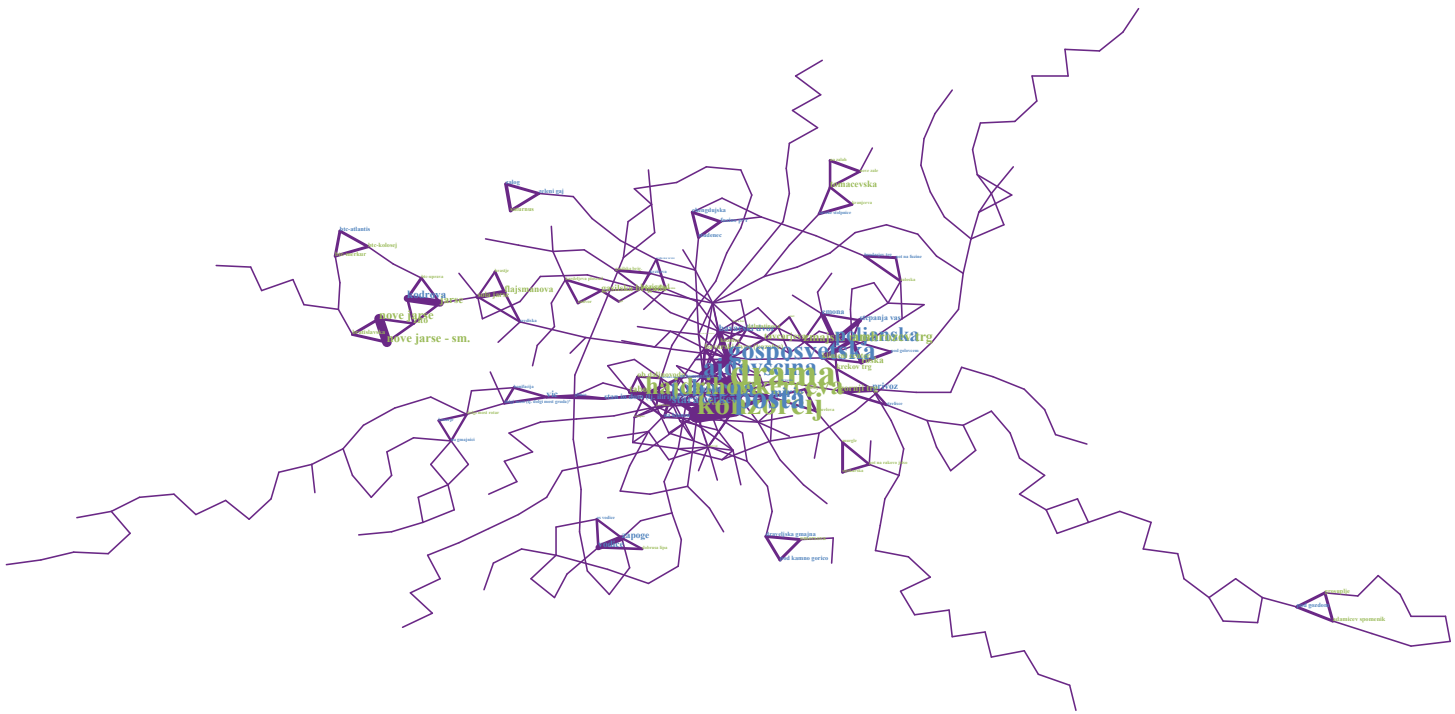


[¶] reduced to largest connected component of simple undirected graph

networkology *μ-embeddedness*

— *μ-corrected embeddedness* $\widetilde{\theta}$ in partial LPP network[‖]

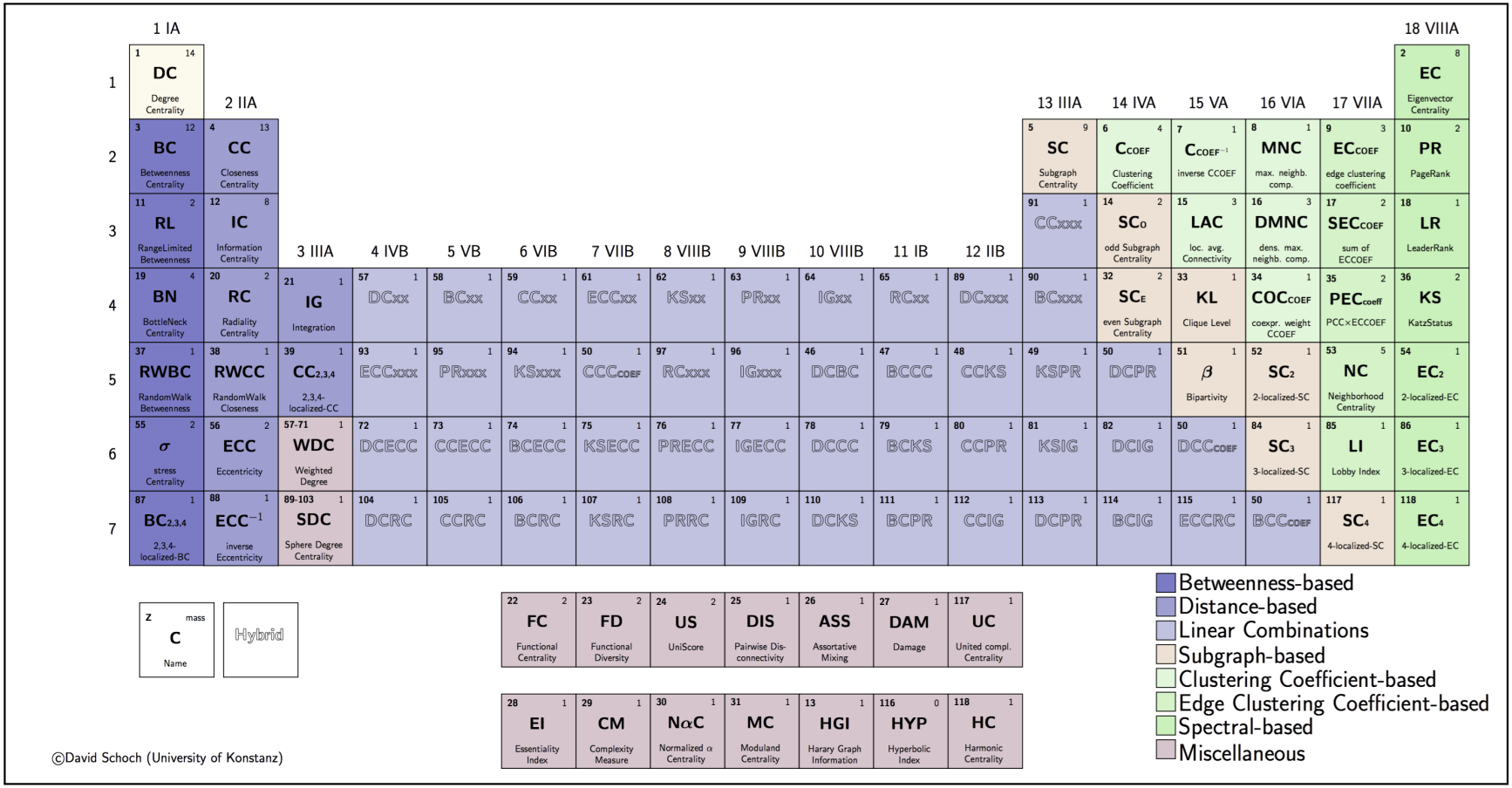
— *highest* $\widetilde{\theta}_{ij} = 0.4$ links are $\{$*Pošta, Konzorcij*$\}$ *etc.*



[‖] reduced to largest connected component of simple undirected graph

bridging *overview*

which *links* are most *important*?



bridging *references*

A.-L. Barabási.
Network Science.
Cambridge University Press, Cambridge, 2016.

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.
Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition.
Cambridge University Press, Cambridge, 2011.

Ernesto Estrada and Philip A. Knight.
A First Course in Network Theory.
Oxford University Press, 2015.

Linton C. Freeman, Stephen P. Borgatti, and Douglas R. White.
Centrality in valued graphs: A measure of betweenness based on network flow.
Soc. Networks, 13(2):141–154, 1991.

L. Freeman.
A set of measures of centrality based on betweenness.
Sociometry, 40(1):35–41, 1977.

Pablo Jensen, Matteo Morini, Marton Karsai, Tommaso Venturini, Alessandro Vespignani, Mathieu Jacomy, Jean-Philippe Cointet, Pierre Merckle, and Eric Fleury.
Detecting global bridges in networks.
J. Complex Netw., 4(3):319–329, 2016.

M. E. J. Newman.
A measure of betweenness centrality based on random walks.
Soc. Networks, 27(1):39–54, 2005.

bridging *references*

Mark E. J. Newman.
Networks: An Introduction.
Oxford University Press, Oxford, 2010.

J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási.
Structure and tie strengths in mobile communication networks.
P. Natl. Acad. Sci. USA, 104(18):7332–7336, 2007.

E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and Albert László Barabási.
Hierarchical organization of modularity in metabolic networks.
Science, 297(5586):1551–1555, 2002.

small-world networks

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2019/20

small-world *documentary*

connected

the power of six degrees

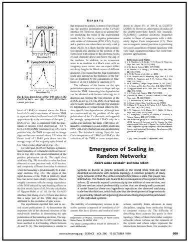
documentary on small-world and scale-free networks


letters to nature

Collective dynamics of 'small-world' networks

Duncan J. Watts & Steven H. Strogatz

[WS98]


Emergence of Scaling in Random Networks

Albert-László Barabási and Réka Albert

[BA99]

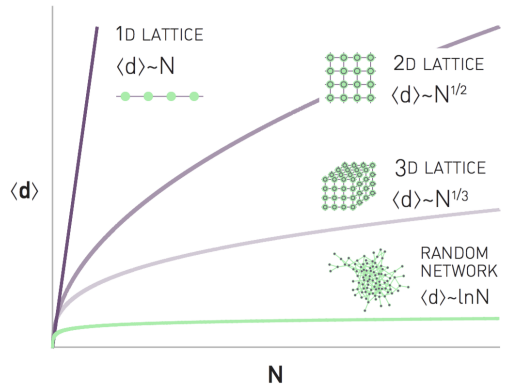

letters to nature

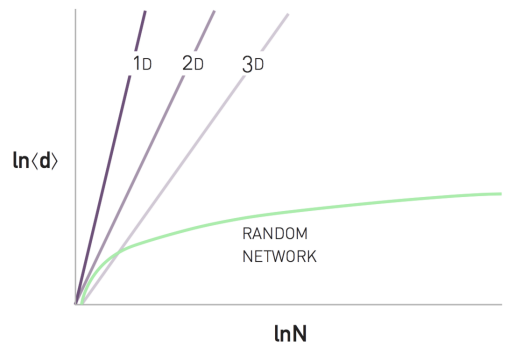
Error and attack tolerance of complex networks

[AJB00]

small-world *phenomenon*

- for *regular lattices*
  - *high clustering* $\langle C \rangle \gg 0$
  - long distances $\langle d \rangle \simeq n^{1/D}$
- in *random graphs* [ER59]
  - low clustering $\langle C \rangle = \frac{\langle k \rangle}{n-1}$
  - *short distances* $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$
- real *small-world networks* [WS98]
  - *high clustering* $\langle C \rangle \gg 0$
  - *short distances* $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$





- $\langle d \rangle = 4.74$ for *Facebook* friendships [BBR+12] while $\frac{\ln n}{\ln \langle k \rangle} = 3.98$
- $\langle C \rangle = 0.61$ for *Facebook* social circles [NL12] while $\rho < 10^{-6}$

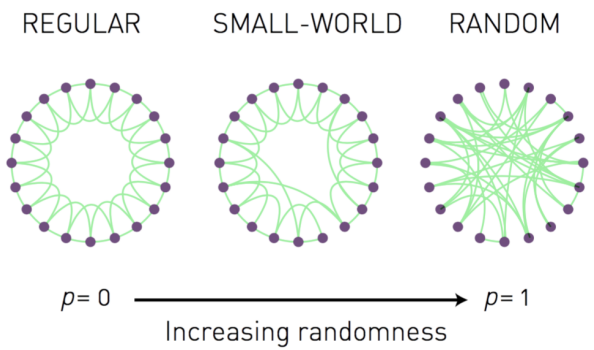
small-world *model*

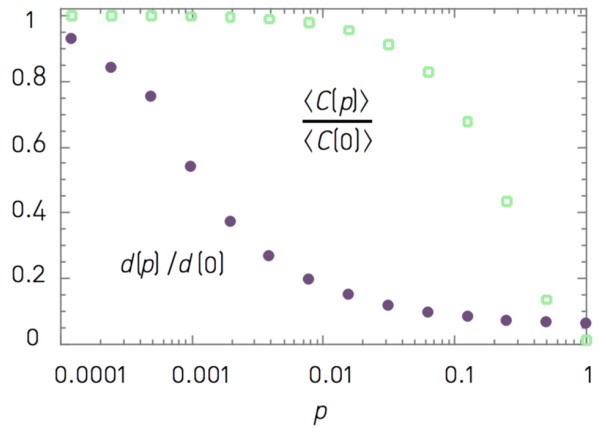
— $G(n, k, p)$ *small-world* model [WS98]

— *randomly rewire* $pnk/2$ *links* of *regular lattice*

— *conceptually interesting* but *practically inapplicable*

- for *some* p *small-world* with $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$ and $\langle C \rangle \gg 0$
- for $p = 1$ *random graph* with $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$
- for $p = 0$ *regular lattice* with $C = \frac{3(k-2)}{4(k-1)}$





see small-world model NetLogo demo

small-world *networks*

— *clustering coefficient* $\langle C \rangle$ in *real small-world* networks

— *average distance* $\langle d \rangle$ in *real small-world* networks

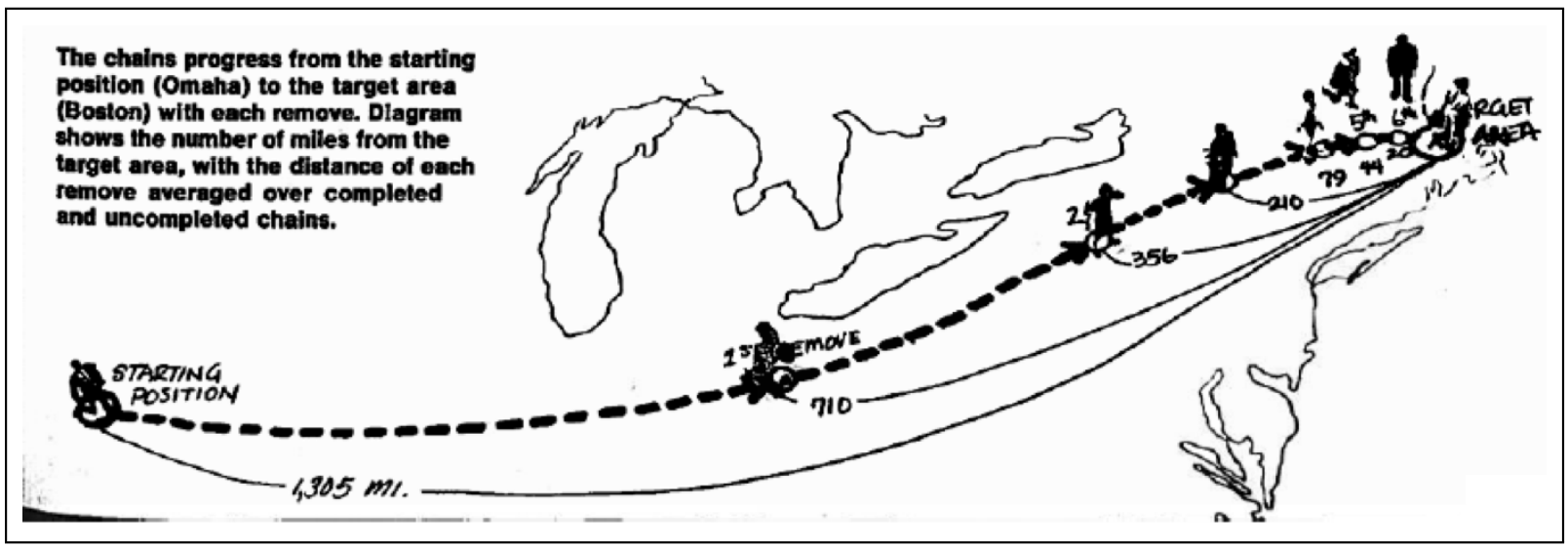
network	n	$\langle C \rangle$	$\gg \frac{\langle k \rangle}{n-1}$	$\langle d \rangle$	$\approx \frac{\ln n}{\ln \langle k \rangle}$
southern women [DGG41]	32	0.000	0.179	2.31	2.02
karate club [Zac77]	34	0.571	0.139	2.41	2.31
American football [GN02]	115	0.403	0.094	2.51	2.00
Java dependencies [ŠB11]	1368	0.497	0.012	2.21	2.59
Facebook circles [ML12]	4039	0.606	0.011	3.69	2.20
physics collaboration [New01]	36 458	0.657	0.000	5.50	4.68
Enron e-mails [LLDM09]	36 692	0.497	0.001	3.39	3.51
Internet map [HJJ+03]	75 885	0.160	0.000	5.83	5.01
actors collaboration [BA99]	382 219	0.780	0.000	≈ 3.6	2.94
physics citation [ŠFB14]	438 943	0.227	0.000	≈ 5.0	4.23
patent citation [HJT01]	3 774 768	0.076	0.000	≈ 8.1	6.98
Facebook snowball [Fer12]	8 217 272	0.019	0.000	≈ 6.8	14.23

small-world *experiments*

— 6 *degrees of separation* in *letter* passing as $\langle d \rangle = 6.2$ [Mil67]

— 4/7 *degrees of separation* in *e-mail* communication [DMW03]

— 4 *degrees of separation* on *Facebook* as $\langle d \rangle = 4.74$ [BBR+12]



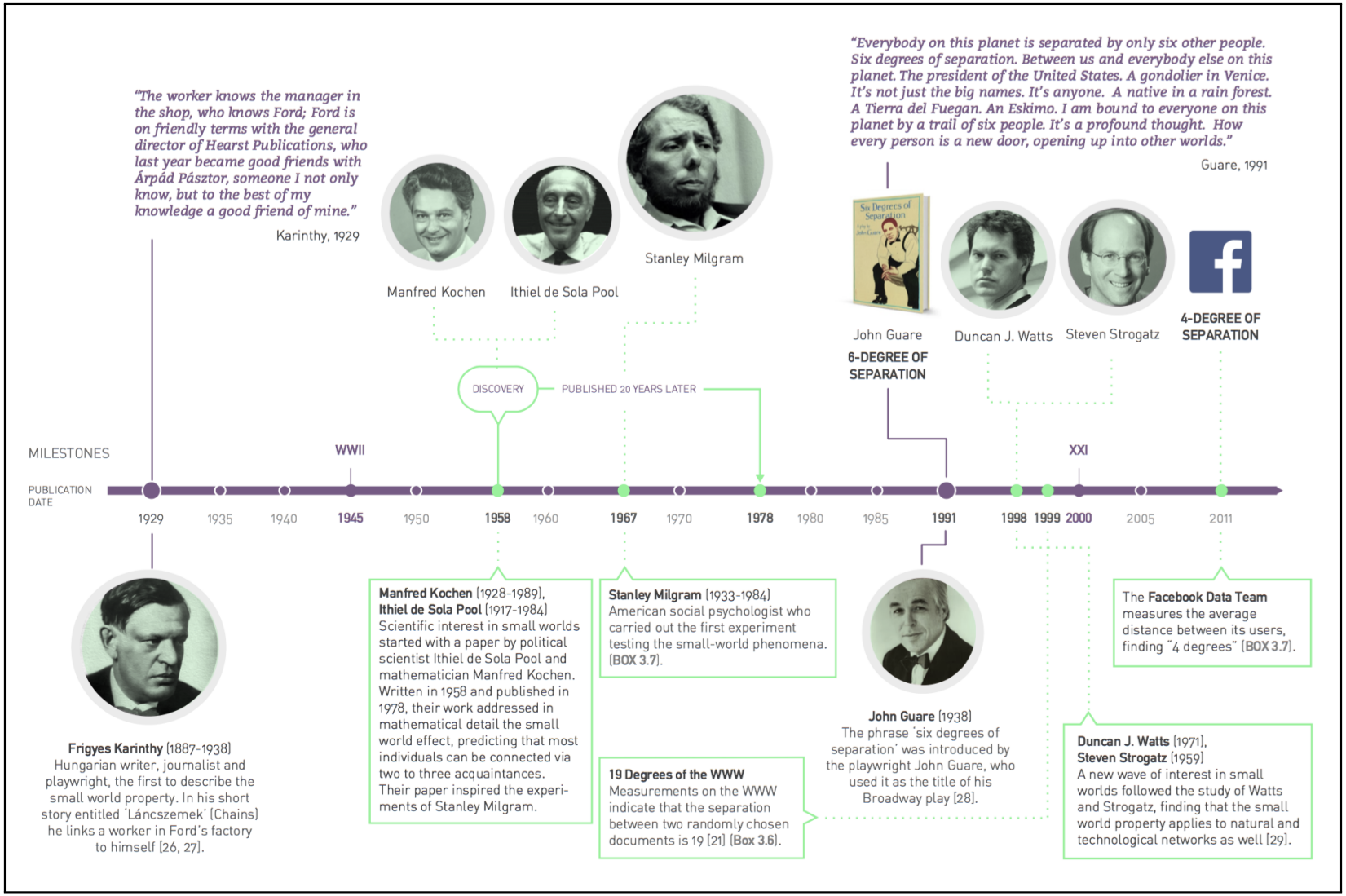
— the *strength* (*weakness*) of *weak* (*strong*) *ties* [Gra73]

small-world *navigation*

does existence of short paths imply

navigable small-world by *decentralized search*? [Kle00]

small-world *history*



small-world *references*

Reka Albert, Hawoong Jeong, and Albert Laszlo Barabasi.
Error and attack tolerance of complex networks.
Nature, 406(6794):378–382, 2000.

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286(5439):509–512, 1999.

A.-L. Barabási.
Network Science.
Cambridge University Press, Cambridge, 2016.

Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna.
Four degrees of separation.
In *Proceedings of the ACM International Conference on Web Science*, pages 45–54, Evanston, IL, USA, 2012.

A. Davis, B. B. Gardner, and M. R. Gardner.
Deep South.
Chicago University Press, Chicago, 1941.

Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts.
An experimental study of search in clobal social networks.
Science, 301(5634):827–829, 2003.

David Easley and Jon Kleinberg.
Networks, Crowds, and Markets: Reasoning About a Highly Connected World.
Cambridge University Press, Cambridge, 2010.

Ernesto Estrada and Philip A. Knight.
A First Course in Network Theory.
Oxford University Press, 2015.

small-world *references*

P. Erdős and A. Rényi.

On random graphs I.
Publ. Math. Debrecen, 6:290–297, 1959.

Stefano Ferretti.

On the degree distribution of faulty peer-to-peer overlays.
ICST Transactions on Complex Systems, 2012.

M. Girvan and M. E. J Newman.

Community structure in social and biological networks.
P. Natl. Acad. Sci. USA, 99(12):7821–7826, 2002.

Mark S. Granovetter.

The strength of weak ties.
Am. J. Sociol., 78(6):1360–1380, 1973.

M Hoerdt, M Jaeger, A James, D Magoni, J Maillard, D Malka, and P Merindol.

Internet {IP}v4 overlay map produced by network cartographer (nec), 2003.

B. H. Hall, A. B. Jaffe, and M. Tratjenberg.

The NBER patent citation data file: Lessons, insights and methodological tools.
Technical report, National Bureau of Economic Research, 2001.

Jon M. Kleinberg.

Navigation in a small world.
Nature, 406(6798):845, 2000.

Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney.

Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters.
Internet Math., 6(1):29–123, 2009.

small-world *references*

Stanley Milgram.

The small world problem.
Psychol. Today, 1(1):60–67, 1967.

Seth A. Myers and Jure Leskovec.

Clash of the contagions: Cooperation and competition in information diffusion.
In *Proceedings of the IEEE International Conference on Data Mining*, 2012.

M. E. J. Newman.

The structure of scientific collaboration networks.
P. Natl. Acad. Sci. USA, 98(2):404–409, 2001.

Mark E. J. Newman.

Networks: An Introduction.
Oxford University Press, Oxford, 2010.

Azree Nazri and Pietro Lio.

Investigating meta-approaches for reconstructing gene networks in a mammalian cellular context.
PLoS ONE, 7(1):e28713, 2012.

Lovro Šubelj and Marko Bajec.

Community structure of complex software systems: Analysis and applications.
Physica A, 390(16):2968–2975, 2011.

Lovro Šubelj, Dalibor Fiala, and Marko Bajec.

Network-based statistical comparison of citation topology of bibliographic databases.
Sci. Rep., 4:6496, 2014.

D. J. Watts and S. H. Strogatz.

Collective dynamics of 'small-world' networks.
Nature, 393(6684):440–442, 1998.

small-world *references*

Wayne W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthropol. Res., 33(4):452–473, 1977.

course *project*

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2019/20

project *overview*

- substantial *network analysis project*
- project *report submitted* (to arXiv.org)
- informal *presentations in front of class*

- project should go *beyond this course*
- project can borrow *from other courses*
- students should *work in pairs*/*triplets*

project *types*

(a) *analytical derivation* of rigorous *theoretical results*

(e) *empirical evaluation* of *methods*, *models* or *networks*

(d) *design* of *novel methods*, *models* or *algorithms*

(i) *scalable implementation* of *methods* or *algorithms*

(c) *contribution* to *Slovenian networks community*

— up to *eight-page report* (to arXiv.org)
  - only for (a), (e) and (d) projects

— *implementation* made *publicly available*
  - only for (i) and (c) projects

— interesting *half-hour* (Cookie) *seminar*
  - only for *very* good projects

project *delivery*

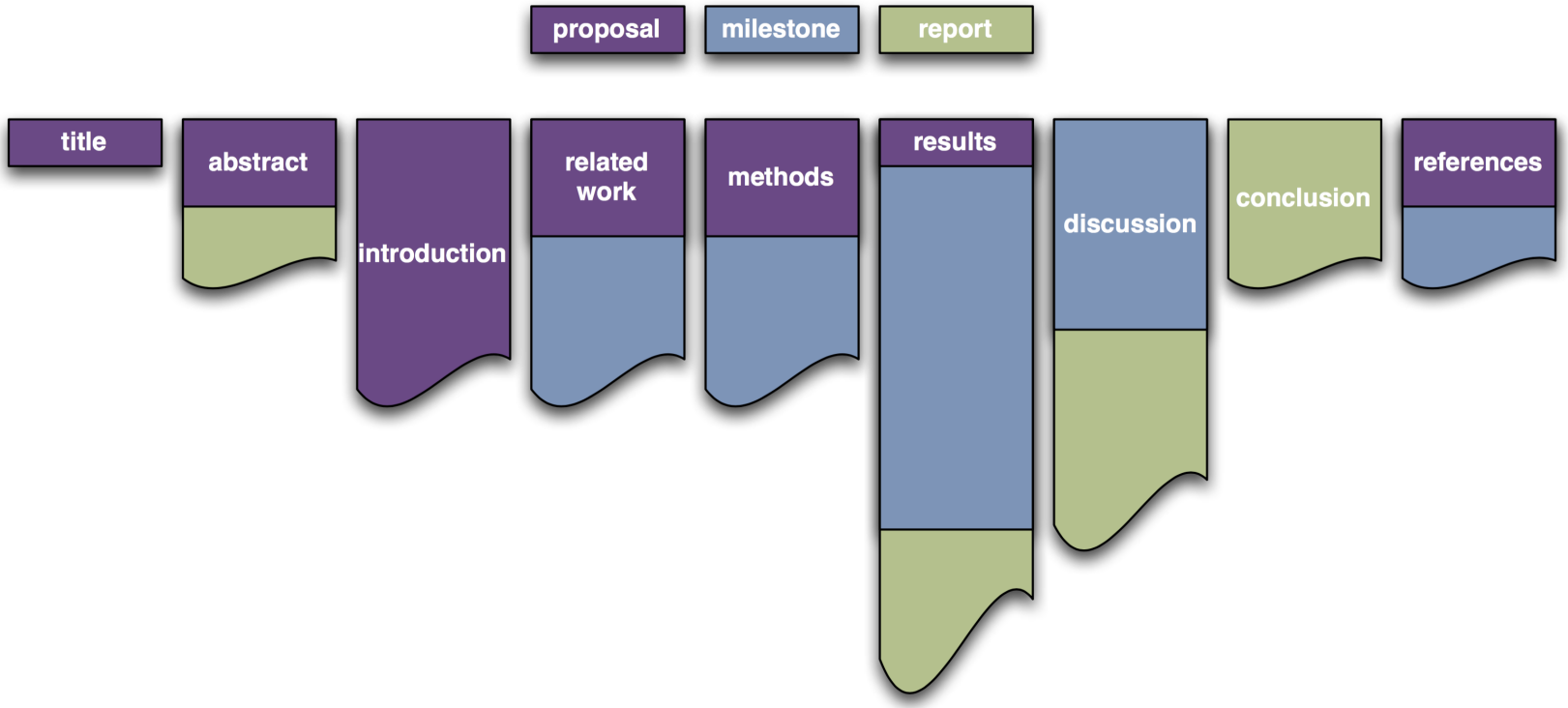
— course *project report* breakdown
  - *project proposal* as *reaction* paper (≈ 1-3 pages)
  - *midterm project milestone* report (≈ 50% done)
  - *project report* as *scientific* paper (≈ 4-8 pages)

— informal *project presentations* to get *feedback*

week	**presentations**	**due**
⋮	⋮	⋮
12	*proposal* (May 4th)	
13		*proposal* (May 11th)
14		
15		*milestone* (May 25th)
16	*project* (Jun 1st)	
17		*project* (Jun 8th)

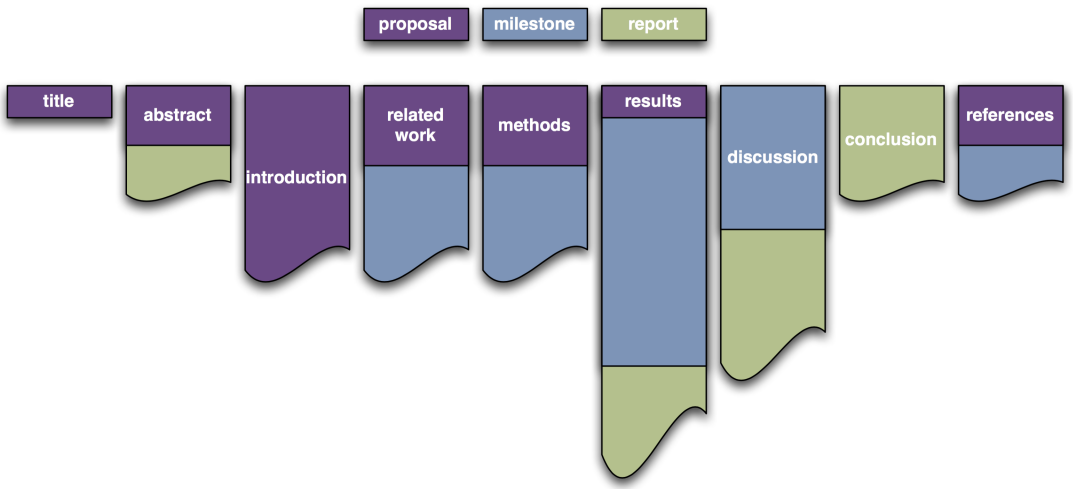
— considering *individual students' contributions*

project *report*



project *proposal*

- *abstract* is brief *summary of problem* and *contribution*
- *introduction* includes *problem motivation* and *background*
- *related work* summarizes selected *relevant literature*
- *methods* describes *used methods* and *techniques*
- *results* already gives some *preliminary results*
- etc.



project *steps*

1. read enough *scientific papers*
2. decide on your *project domain*
3. find *relevant data* and check it
4. decide on your *project problem*
5. clearly state your *project hypotheses*
6. select *appropriate* tools and *techniques*
7. conduct *preliminary analysis* of hypotheses
8. find *related work* papers and read them

x. *analyze* your hypotheses

y. *refine* your hypotheses

z. *iterate* until deadline

project *tips*

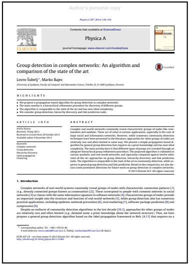
- know your *project domain*
- assume *data you can* actually *get*
- make *hypotheses* as *simple* as possible
- use *simplest methods* and *techniques* possible
- make *results* practically *useful* and *valuable*
- make project *completion feasible*
- *work* and *write* simultaneously
- *start* working on project *early*

project *templates*

- standard *Elsevier journals* manuscripts
- LaTeX *template* and example *manuscript*

introduction

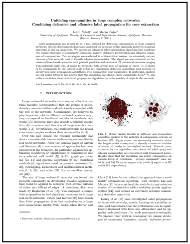
conclusions



proofs

project *templates*

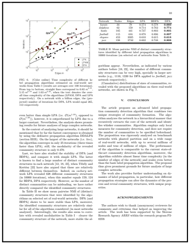
- *American Physical Society* manuscripts
- LaTeX *template* and example *manuscript*



Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction

I. INTRODUCTION

introduction



V. CONCLUSION

ACKNOWLEDGMENTS

conclusions

PHYSICAL REVIEW E 83, 036103 (2011)

Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction

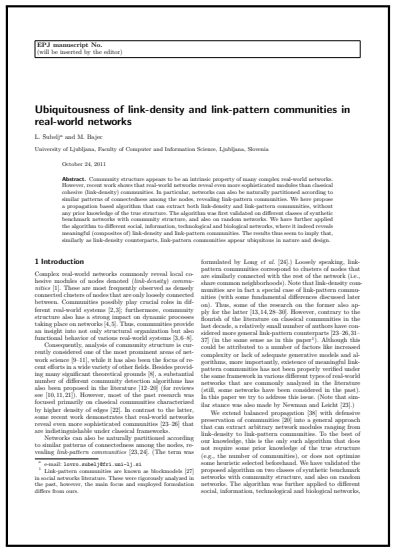
I. INTRODUCTION

proofs

project *templates*

— *European Physical Society* manuscripts

— LaTeX *template* and example *manuscript*



EPJ manuscript No.
(will be inserted by the editor)

Ubiquitousness of link-density and link-pattern communities in real-world networks

L. Šubelj* and M. Bajec

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia

October 24, 2011

Abstract. Community structure appears to be an intrinsic property of many complex real-world networks. However, recent work shows that real-world networks reveal even more sophisticated modules than classical cohesive (link-density) communities. In particular, networks can also be naturally partitioned according to similar patterns of connectedness among the nodes, revealing link-pattern communities. We here propose a propagation based algorithm that can extract both link-density and link-pattern communities, without any prior knowledge of the true structure. The algorithm was first validated on different classes of synthetic benchmark networks with community structure, and also on random networks. We have further applied the algorithm to different social, information, technological and biological networks, where it indeed reveals meaningful (composites of) link-density and link-pattern communities. The results thus seem to imply that, similarly as link-density counterparts, link-pattern communities appear ubiquitous in nature and design.

1 Introduction

Complex real-world networks commonly reveal local cohesive modules of nodes denoted (*link-density*) communities [1]. These are most frequently observed as densely connected clusters of nodes that are only loosely connected between. Communities possibly play crucial roles in different real-world systems [2,3]; furthermore, community structure also has a strong impact on dynamic processes taking place on networks [4,5]. Thus, communities provide an insight into not only structural organization but also functional behavior of various real-world systems [3,6–8].

Consequently, analysis of community structure is currently considered one of the most prominent areas of network science [9–11], while it has also been the focus of recent efforts in a wide variety of other fields. Besides providing many significant theoretical grounds [8], a substantial number of different community detection algorithms has also been proposed in the literature [12–20] (for reviews see [10,11,21]). However, most of the past research was focused primarily on classical communities characterized by higher density of edges [22]. In contrast to the latter, some recent work demonstrates that real-world networks reveal even more sophisticated communities [23–26] that are indistinguishable under classical frameworks.

Networks can also be naturally partitioned according to similar patterns of connectedness among the nodes, revealing *link-pattern communities* [23,24]. (The term was formulated by Long *et al.* [24].) Loosely speaking, link-pattern communities correspond to clusters of nodes that are similarly connected with the rest of the network (i.e., share common neighborhoods). Note that link-density communities are in fact a special case of link-pattern communities (with some fundamental differences discussed later on). Thus, some of the research on the former also apply for the latter [13,14,28–30]. However, contrary to the flourish of the literature on classical communities in the last decade, a relatively small number of authors have considered more general link-pattern counterparts [23–26,31–37] (in the same sense as in this paper[1]). Although this could be attributed to a number of factors like increased complexity or lack of adequate generative models and algorithms, more importantly, existence of meaningful link-pattern communities has not been properly verified under the same framework in various different types of real-world networks that are commonly analyzed in the literature (still, some networks have been considered in the past). In this paper we try to address this issue. (Note that similar stance was also made by Newman and Leicht [23].)

We extend balanced propagation [38] with defensive preservation of communities [26] into a general approach that can extract arbitrary network modules ranging from link-density to link-pattern communities. To the best of our knowledge, this is the only such algorithm that does not require some prior knowledge of the true structure (e.g., the number of communities), or does not optimize some heuristic selected beforehand. We have validated the proposed algorithm on two classes of synthetic benchmark networks with community structure, and also on random networks. The algorithm was further applied to different social, information, technological and biological networks,

* e-mail: lovro.subelj@fri.uni-lj.si

[1] Link-pattern communities are known as blockmodels [27] in social networks literature. These were rigorously analyzed in the past, however, the main focus and employed formulation differs from ours.

introduction

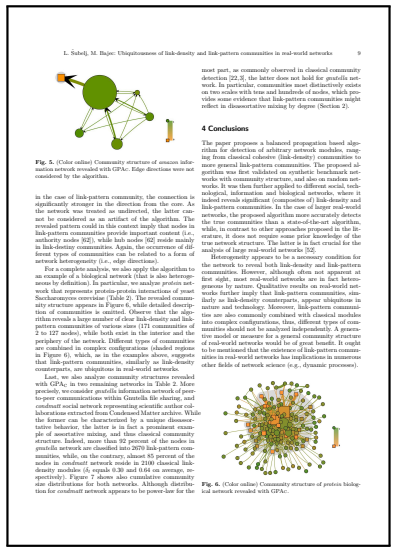


Fig. 5. (Color online) Community structure of *smallcon* information network revealed with GPAc. Edge directions were not considered by the algorithm.

in the case of link-pattern community, the connection is significantly stronger in the direction from the core. As the network was treated as undirected, the latter cannot be considered as an artifact of the algorithm. The revealed pattern could in this context imply that nodes in link-pattern communities provide important content (i.e., authority nodes [62]), while hub nodes [62] reside mainly in link-destiny communities. Again, the occurrence of different types of communities can be related to a form of network heterogeneity (i.e., edge directions).

For a complete analysis, we also apply the algorithm to an example of a biological network (that is also heterogeneous by definition). In particular, we analyze *protein* network that represents protein-protein interactions of yeast Saccharomyces cerevisiae (Table 2). The revealed community structure appears in Figure 6, while detailed description of communities is omitted. Observe that the algorithm reveals a large number of clear link-density and link-pattern communities of various sizes (17 communities of 2 to 127 nodes), while both exist in the interior and the periphery of the network. Different types of communities are combined in complex configurations (shaded regions in Figure 6), which, as in the examples above, suggests that link-pattern communities, similarly as link-density counterparts, are ubiquitous in real-world networks.

Last, we also analyze community structures revealed with GPAc in two remaining networks in Table 2. More precisely, we consider *gnutella* information network of peer-to-peer communications within Gnutella file sharing, and *condmat* social network representing scientific author collaborations extracted from Condensed Matter archive. While the former can be characterized by a unique dissociative behavior, the latter is in fact a prominent example of assortative mixing, and thus classical community structure. Indeed, more than 92 percent of the nodes in *gnutella* network are classified into 2670 link-pattern communities, while, on the contrary, almost 85 percent of the nodes in *condmat* network reside in 2100 classical link-density modules (Δ equals 0.30 and 0.64 on average, respectively). Figure 7 shows also cumulative community size distributions for both networks. Although distribution for *condmat* network appears to be power-law for the most part, as commonly observed in classical community detection [22,3], the latter does not hold for *gnutella* network. In particular, communities most distinctively exists on two scales with tens and hundreds of nodes, which provides some evidence that link-pattern communities might reflect in dissassortative mixing by degree (Section 2).

4 Conclusions

The paper proposes a balanced propagation based algorithm for detection of arbitrary network modules, ranging from classical cohesive (link-density) communities to more general link-pattern communities. The proposed algorithm was first validated on synthetic benchmark networks with community structure, and also on random networks. It was then further applied to different social, technological, information and biological networks, where it indeed reveals significant (composites of) link-density and link-pattern communities. In the case of larger real-world networks, the proposed algorithm more accurately detects the true communities than a state-of-the-art algorithm, while, in contrast to other approaches proposed in the literature, it does not require some prior knowledge of the true network structure. The latter is in fact crucial for the analysis of large real-world networks [52].

Heterogeneity appears to be a necessary condition for the network to reveal both link-density and link-pattern communities. However, although often not apparent at first sight, most real-world networks are in fact heterogeneous by nature. Qualitative results on real-world networks further imply that link-pattern communities, similarly as link-density counterparts, appear ubiquitous in nature and technology. Moreover, link-pattern communities are also commonly combined with classical modules into complex configurations, thus, different types of communities should not be analyzed independently. A generative model or measure for a general community structure of real-world networks would be of great benefit. It ought to be mentioned that the existence of link-pattern communities in real-world networks has implications in numerous other fields of network science (e.g., dynamic processes).

Fig. 6. (Color online) Community structure of protein biological network revealed with GPAc.

conclusions

Eur. Phys. J. B (2012) 85: 32
DOI: 10.1140/epjb/e2011-20468-7

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Ubiquitousness of link-density and link-pattern communities in real-world networks

L. Šubelj* and M. Bajec

University of Ljubljana, Faculty of Computer and Information Science, 1000 Ljubljana, Slovenia

Received 9 June 2011 / Received in final form 24 October 2011
Published online 18 January 2012 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2012

Abstract. Community structure appears to be an intrinsic property of many complex real-world networks. However, recent work shows that real-world networks reveal even more sophisticated modules than classical cohesive (link-density) communities. In particular, networks can also be naturally partitioned according to similar patterns of connectedness among the nodes, revealing link-pattern communities. We here propose a propagation based algorithm that can extract both link-density and link-pattern communities, without any prior knowledge of the true structure. The algorithm was first validated on different classes of synthetic benchmark networks with community structure, and also on random networks. We have further applied the algorithm to different social, information, technological and biological networks, where it indeed reveals meaningful (composites of) link-density and link-pattern communities. The results thus seem to imply that, similarly as link-density counterparts, link-pattern communities appear ubiquitous in nature and design.

1 Introduction

Complex real-world networks commonly reveal local cohesive modules of nodes denoted (*link-density*) communities [1]. These are most frequently observed as densely connected clusters of nodes that are only loosely connected between. Communities possibly play crucial roles in different real-world systems [2,3]; furthermore, community structure also has a strong impact on dynamic processes taking place on networks [4,5]. Thus, communities provide an insight into not only structural organization but also functional behavior of various real-world systems [3,6–8].

Consequently, analysis of community structure is currently considered one of the most prominent areas of network science [9–11], while it has also been the focus of recent efforts in a wide variety of other fields. Besides providing many significant theoretical grounds [8], a substantial number of different community detection algorithms has also been proposed in the literature [12–20] (for reviews see [10,11,21]). However, most of the past research was focused primarily on classical communities characterized by higher density of edges [22]. In contrast to the latter, some recent work demonstrates that real-world networks reveal even more sophisticated communities [23–26] that are indistinguishable under classical frameworks.

Networks can also be naturally partitioned according to similar patterns of connectedness among the nodes, revealing *link-pattern communities* [23,24]. (The term was formulated by Long et al. [24]. Loosely speaking, link-pattern communities correspond to clusters of nodes that are similarly connected with the rest of the network (i.e., share common neighborhoods). Note that link-density communities are in fact a special case of link-pattern communities (with some fundamental differences discussed later on). Thus, some of the research on the former also apply for the latter [13,14,28–30]. However, contrary to the flourish of the literature on classical communities in the last decade, a relatively small number of authors have considered more general link-pattern counterparts [23–26,31–37] (in the same sense as in this paper[1]). Although this could be attributed to a number of factors like increased complexity or lack of adequate generative models and algorithms, more importantly, existence of meaningful link-pattern communities has not been properly verified under the same framework in various different types of real-world networks that are commonly analyzed in the literature (still, some networks have been considered in the past). In this paper we try to address this issue. (Note that similar stance was also made by Newman and Leicht [23].)

We extend balanced propagation [38] with defensive preservation of communities [26] into a general approach that can extract arbitrary network modules ranging from link-density to link-pattern communities. To the best of our knowledge, this is the only such algorithm that does not require some prior knowledge of the true structure (e.g., the number of communities), or does not optimize

* e-mail: lovro.subelj@fri.uni-lj.si

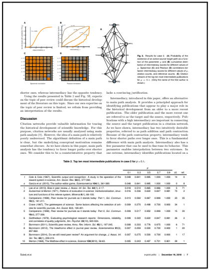
Link-pattern communities are known as blockmodels [27] in social networks literature. These were rigorously analyzed in the past, however, the main focus and employed formulation differs from ours.

proofs

project *templates*

- suggested top-notch *PNAS* manuscripts
- LaTeX *template* and example *manuscript*

introduction



conclusions

proofs