

Odkrivanje skupin vozlišč v omrežjih z izmenjavo oznak

Lovro Šubelj

Univerza v Ljubljani,
Fakulteta za računalništvo in informatiko

FIS, 20.1.2014

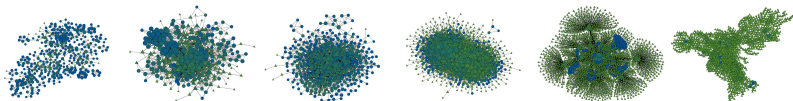
Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Analiza omrežij

Številne vrste realnih omrežij: Newman (2003)

socialna, informacijska, tehnološka, biološka itd.



Analiza omrežij: Newman (2008)

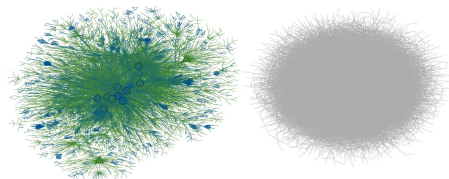
- preučevanje zgradbe realnih omrežij
- razvoj algoritmov in pristopov
- primeri uporabe v praksi

Izjemno aktivno področje v številnih znanostih:

matematika, fizika, računalništvo, družboslovje, biologija itd.

Zgradba realnih omrežij

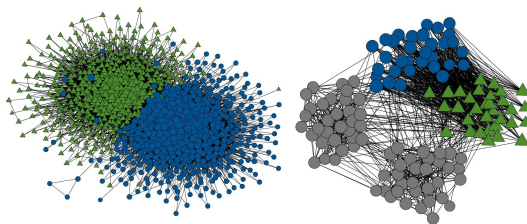
- lastnosti omrežij (npr. mali svet, brezlestvičnost)
- večje skupine vozlišč (npr. skupnosti, moduli)
- manjši vzorci vozlišč (npr. motivi, grafki)
- lastnosti vozlišč (npr. kazala, viri)
- povezave, procesi ipd.



Omejimo se na neusmerjena omrežja.

Karakteristične skupine vozlišč

- skupnosti (*community*) Girvan and Newman (2002)
(povezane) skupine tesno povezanih vozlišč
- moduli (*module*) Newman and Leicht (2007)
(nepovezane) skupine podobno povezanih vozlišč



Omejimo se na neprekrivajoče skupine vozlišč.

Vsebina

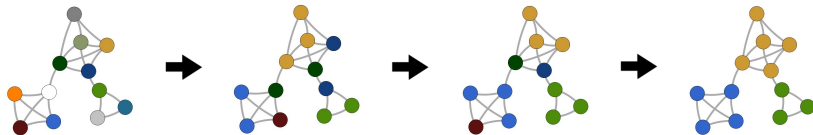
- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Osnovna izmenjava oznak

Osnovna izmenjava oznak (*label propagation*): Raghavan et al. (2007)

$$g_i = \operatorname{argmax}_g \sum_{v_j \in N_i} \delta(g_j, g)$$

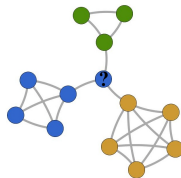
g_i posodablamo v naključnem vrstnem redu.



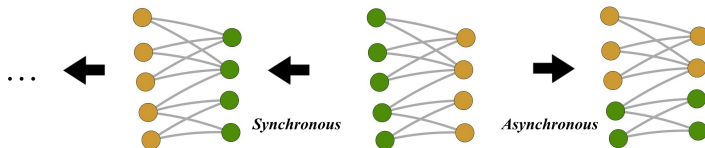
g_i je oznaka skupine vozlišča v_i in δ je Kroneckerjev delta.

Nedeterminističnost izmenjave oznak

v_i ohrani oznako g_i v kolikor je med najpogostejšimi. Barber and Clark (2009)



Oznake g_i posodabljamo zaporedno (*asynchronous*). Raghavan et al. (2007)



Analiza izmenjave oznak

Prednosti:

- brez predhodnega znanja (npr. število skupin)
- skoraj linearna časovna zahtevnost
- enostavna implementacija

Slabosti:

robustnost več razvrstitev že v manjših omrežjih Tibély and Kertész (2008)

natančnost slabša natančnost v omrežjih z nejasno zgradbo Leung et al. (2009)

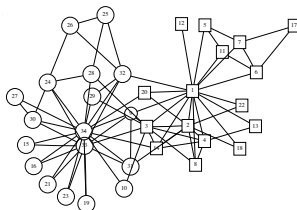
splošnost pristop omejen na odkrivanje skupnosti Šubelj and Bajec (2012b)

Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč
 - Osnovna izmenjava oznak
 - **Uravnotežena izmenjava oznak**
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Robustnost izmenjave oznak

Osnovna izmenjava vrne > 500 razvrstitev v skupine. Tibély and Kertész (2008)

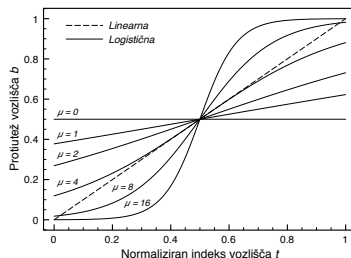


Vrstni red obravnave g_i se odraža kot preference vozlišč f_i . Šubelj and Bajec (2011a)

$$g_i = \operatorname{argmax}_g \sum_{v_j \in N_i} f_j \cdot \delta(g_j, g)$$

f_i je “moč” širjenja oznake vozlišča v_i . Leung et al. (2009)

Uravnotežena izmenjava oznak



Uravnotežena izmenjava (*balanced propagation*): Šubelj and Bajec (2011a)

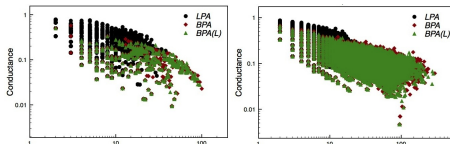
$$g_i = \operatorname{argmax}_g \sum_{v_j \in N_i} b_j \cdot \delta(g_j, g) \quad b_i = \frac{1}{1 + e^{-\mu(t_i - \lambda)}}$$

b_i je protiutež in $t_i \in (0, 1]$ normaliziran indeks vozlišča v_i .

Eksperimentalni rezultati

različnih razvrstitev v 1000 ponovitvah: Šubelj and Bajec (2011a)

	<i>karate</i>	<i>dolphins</i>	<i>books</i>	<i>football</i>	<i>jazz</i>	<i>elegans</i>
Osnovna izmenjava	184	525	269	414	63	707
Uravnotežena izmenjava	19	36	29	154	20	75



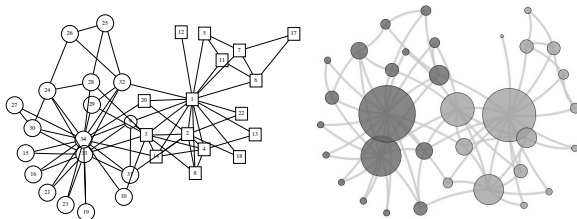
Robustnost se izboljša na račun časovne zahtevnosti. Šubelj and Bajec (2011c)

Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč**
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak**
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

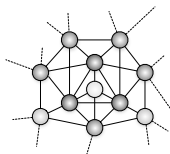
Natančnost izmenjave oznak

Natančnost izboljšamo z uporabo preferenc vozlišč f_i . Leung et al. (2009)

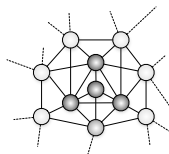


Osnovne lastnosti vozlišč niso primerne za preference f_i . Šubelj and Bajec (2011b)

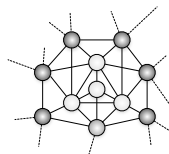
Zadržana in napadalna izmenjava oznak



Stopnje vozlišč



Zadržana



Napadalna

Zadržana izmenjava (*defensive propagation*): Šubelj and Bajec (2011b)

$$g_i = \operatorname{argmax}_g \sum_{v_j \in N_i} p_j \cdot \delta(g_j, g)$$

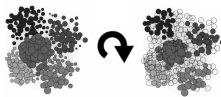
Napadalna izmenjava (*offensive propagation*):

$$g_i = \operatorname{argmax}_g \sum_{v_j \in N_i} (1 - p_j) \cdot \delta(g_j, g)$$

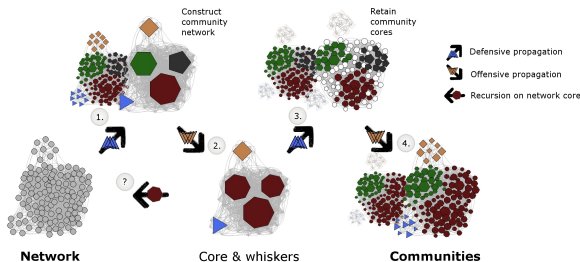
p_i je verjetnost naključnega sprehajalca znotraj skupine g_i .

Napredna izmenjava oznak

Zadržana (napadalna) izmenjava doseže visok priklic (natančnost).



Napredna izmenjava (*diffusion propagation*): Šubelj and Bajec (2011b)



Eksperimentalni rezultati

Natančnost je primerljiva z najboljšimi pristopi. Šubelj and Bajec (2010)

Network	Description	Nodes	Edges	GMO	LPA	LPAD	LPAQ	LPAM	BDPA	DPA	No. CE ^c	T ^c
<i>karate</i>	Zachary's karate club [33]	34	78	0.381	0.416	0.402	0.399	0.420	0.419	0.420	0.02	
<i>dolphins</i>	Lusseau's bottlenose dolphins [38]	62	159		0.529	0.526	0.516	0.529	0.528	0.529	0.59	
<i>books</i>	Co-purchased political books [39]	105	441		0.526	0.519	0.522	0.527	0.527	0.527	0.46	
<i>football</i>	American football league [6]	115	616	0.556	0.606	0.606	0.604	0.605	0.606	0.606	0.37	
<i>elegans</i>	Metabolic network <i>C. elegans</i> [37]	453	2025	0.412	0.421	0.413	0.409	<i>0.452</i>	0.424	0.427^b	0.17	
<i>jazz</i>	Jazz musicians [40]	198	2742	0.439	0.443	0.443	0.445	0.445	0.444	0.444	0.00	
<i>netsci</i>	Network scientists [9]	1589	2742		0.902	0.947			0.907	0.960	1.00	
<i>yeast</i>	Yeast protein interactions [41]	2114	4480		0.694	0.799			0.725	0.824	1.04	
<i>emails</i>	Emails within a university [42]	1133	5451	0.503	0.557	0.560	0.537	<i>0.582</i>	0.555	0.562	0.01	
<i>power</i>	Western US power grid [32]	4941	6594		0.612	0.804			0.668	0.908	1.14	
<i>blogs</i>	Weblogs on politics [43]	1490	16718		0.426	0.426			0.426	0.426	1.00	
<i>pgp</i>	PGP web of trust [44]	10680	24340	0.849	0.754	0.844	0.726	<i>0.884</i>	0.782	0.869	1.08	
<i>asi</i>	Autonomous syst. of Internet [25]	22963	48436		0.511	0.591			0.528	0.600^b	1.02	0 s
<i>codmat³</i>	Cond. Matt. archive 2003 ^a [45]	27519	116181	0.661	0.616	0.683	0.582	<i>0.755</i>	0.634	0.735	1.00	1.5 s
<i>codmat⁵</i>	Cond. Matt. archive 2005 ^a [45]	36458	171736		0.586	0.643			0.608	0.683	1.00	
<i>kdd³</i>	KDD-Cup 2003 dataset [46]	27770	352285		0.624	0.630			0.619	0.617	1.00	3 s
<i>nec</i>	nec web overlay map [47]	75885	357317		0.693	0.738			0.703	0.767	1.03	
<i>epinions</i>	Epinions web of trust [48]	75879	508837		0.382	0.362			0.399	0.402	1.00	4.5 s
<i>amazon³</i>	Amazon co-purchasing 2003 [49]	262111	1.2M		0.682	0.749			0.701	0.857	1.01	20 s
<i>ndedu</i>	Webpages in nd.edu domain [50]	325729	1.5M		0.840	0.890			0.863	0.903	1.14	
<i>google</i>	Web graph of Google [3]	875713	4.3M		0.805	0.923			0.822	0.968	1.01	2.5 m
<i>nber</i>	NBER patents citations [51]	3.8M	16.5M		0.573	0.624			0.583	0.759	1.20	
<i>live</i>	Live Journal friendships [3]	4.8M	69.0M		0.538	0.539			0.557	0.693	1.00	44 m

Časovna zahtevnost skoraj linearna $\mathcal{O}(m^{1,19})$. Šubelj and Bajec (2011b)

m je število povezav v omrežju.

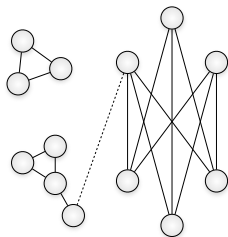
Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 **Odkrivanje skupin vozlišč**
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - **Posplošena izmenjava oznak**
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

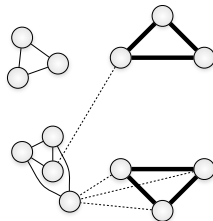
Splošnost izmenjave oznak

Pristop omejen na tesno povezana vozlišča (tj. skupnosti). Šubelj and Bajec (2012b)

Analogija med skupinami:



2 skupnosti, 2 modula



4 skupnosti

Oznake naj si izmenjujejo tudi vozlišča na razdalji dva.

Posplošena izmenjava oznak

Posplošena izmenjava (*general propagation*): Šubelj and Bajec (2012b)

$$g_i = \operatorname{argmax}_g \left(\nu_g \sum_{v_j \in N_i} \delta(g_j, g) + (1 - \nu_g) \sum_{v_j \in N_i} \sum_{v_k \in N_j} \delta(g_k, g) \right)$$

ν_g je blizu ena (nič) za skupnosti (module):

- lastnosti skupin Šubelj and Bajec (2012b)
- nakopičenost vozlišč Šubelj and Bajec (2011c)
- popravljena nakopičenost Šubelj and Bajec (2014)

$\nu_g \in [0, 1]$ je parameter skupine vozlišč g .

Natančnost je vsaj primerljiva z najboljšimi pristopi. Šubelj and Bajec (2014)



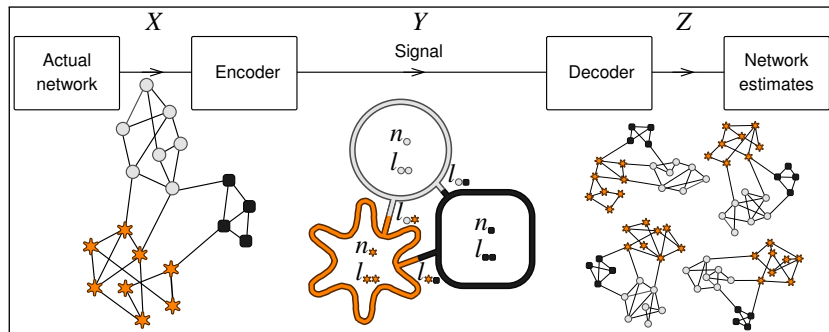
Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč**
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Algoritem *Infomod*

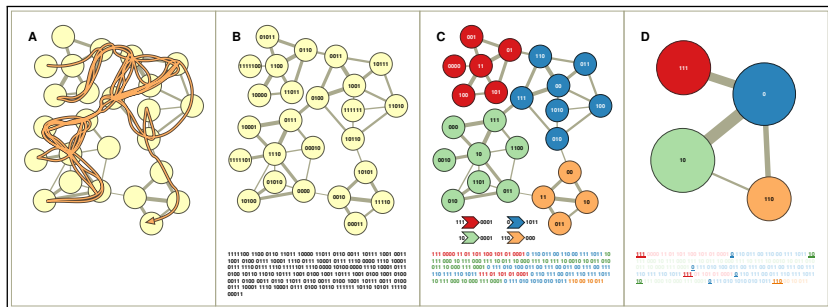
Rosvall and Bergstrom (2007)

Odkrivanje skupin s kodiranjem in metodo največjega verjetja.



Algoritem *Infomap* Rosvall and Bergstrom (2008)

Odkrivanje skupin z naključnimi sprehodi in Huffmanovim kodiranjem.



Ekstrakcija skupin vozlišč Šubelj et al. (2013a)

Group extraction for real-world networks

Lovro Šubelj¹, Neli Blagus & Marko Bajec

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

Background

Complex real-world networks contain characteristic groups of nodes with common linking pattern like densely linked **communities** [1]. These were the focus of most recent work and have diverse applications. However, many real-world networks also contain **other groups of nodes** that can be **overlapping** and other, whereas some parts of the networks reveal **no significant groups**.

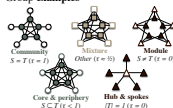
Group formalism

Let S be a group of nodes, T the linking pattern and t the group parameter.

$$\tau(S, T) = \frac{|S \cap T|}{|S \cup T|}$$



Group examples



Group criterion

Let W be the group criterion, L the number of links and μ the (harmonic) mean size.

$$W(S, T) = \mu(S, T) (1 - \mu(S, T)) \left(\frac{L(S, T)}{|S|T|} - \frac{L(S, T^2)}{|S|T|^2} \right)$$

W is a local asymmetric criterion that favors the links between S and T , and penalizes for the links between S and T^2 . (Note, however, that W disregards the links with both endpoints in S^2) For $S = T$, W is consistent with a wide class of other models (e.g., stochastic blockmodel). [2]

Group extraction

A sequential extraction [2] of groups that can be **overlapping**, **nested** etc.

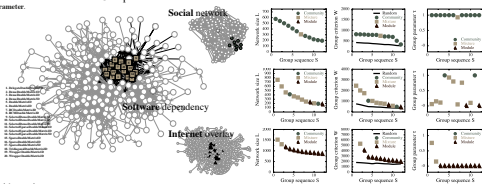
- (1) Find S and T that optimize criterion W (e.g., naive search).
- (2) Extract only the explained links between S and T (and isolated nodes).
- (3) Repeat until W is larger than expected in a random graph (by simulation).

Contributions

1. A simple formalism and criterion for general groups of nodes.
2. An adequate extraction procedure for statistically significant groups.
3. Characterization of the group structure of different real-world networks.

What are characteristic groups of nodes in real-world networks? Network (type) dependent.
What portion of network links is explained by the group structure? Between 60% and 90%.
What portion of network nodes is included in the group structure? More than 50%.

Groups in real-world networks



Network	Nodes	Links	Group	#	S	Community	Core	Mixture	Module	Background
Author collaboration	20	1589	2742	140	1.0	0.94	71% (47%)	0% (0%)	0% (0%)	1% (1%)
American football	[1]	115	613	13	8.6	0.88	59% (83%)	9% (12%)	3% (7%)	0% (0%)
Lucene search engine	[1]	1657	6808	123	12.1	0.55	19% (25%)	1% (2%)	30% (24%)	38% (34%)
Colt computing	[4]	227	963	15	10.3	0.41	7% (11%)	0% (0%)	69% (49%)	4% (6%)
Word adjacency	[3]	112	425	4	11.2	0.28	0% (0%)	0% (0%)	34% (33%)	22% (13%)
Internet overlay	[5]	767	1857	33	10.6	0.08	0% (1%)	12% (4%)	13% (7%)	34% (35%)
Southern women	[6]	32	89	2	4.3	0.00	0% (0%)	0% (0%)	0% (0%)	80% (41%)

All extracted groups are statistically significant at 1% level.

References

- [1] Girvan, M. & Newman, M.E.J.: Community structure in social and biological networks. *P. Natl. Acad. Sci. USA* 99(12), 7821–7826 (2002).
- [2] Zhou, Y., Lorenz, E., & Zhu, J.: Community structure in social networks. *P. Natl. Acad. Sci. USA* 108(18), 7521–7526 (2011).
- [3] Newman, M.E.J.: Finding community structure in networks using the eigenvector method. *Phys. Rev. E* 70(6), 066113 (2004).
- [4] Šubelj, L. & Bajec, M.: Community structure of complex software systems: Analysis and applications. *Phrasa & PPS*, 298–297 (2011).
- [5] Lichtenberg, S., Manning, B., & Finkler, C.: Grouping data: Social networks, social media, and public opinion. In: *Proceedings of the ACM SIGMOD International Conference on Knowledge Discovery and Data Mining (Chicago, IL, USA, 2005)*, pp. 177–187.
- [6] Davis, A., Gardner, B., & Gardner, M.R.: *Deep South* (Chicago University Press, Chicago, 1941).



*Corresponding author: lovro.subelj@fri.uni-lj.si

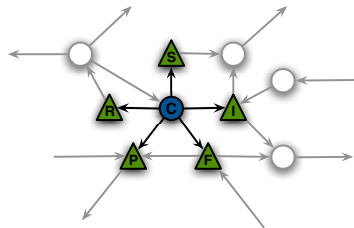
Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Programska omrežja

Omrežja odvisnosti med razredi (*class dependency*): Šubelj and Bajec (2011d)

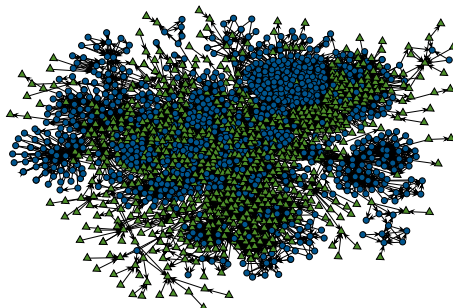
```
class C extends S implements I {  
    F field;  
    public C() { ... }  
    void foo(P parameter) { ... }  
    private R bar() { ... }  
}
```



Podobne splošne lastnosti kot druga realna omrežja. Valverde et al. (2002)

Analiza skupnosti vozlišč

Programska omrežja vsebujejo jasne skupnosti. Šubelj and Bajec (2011d)

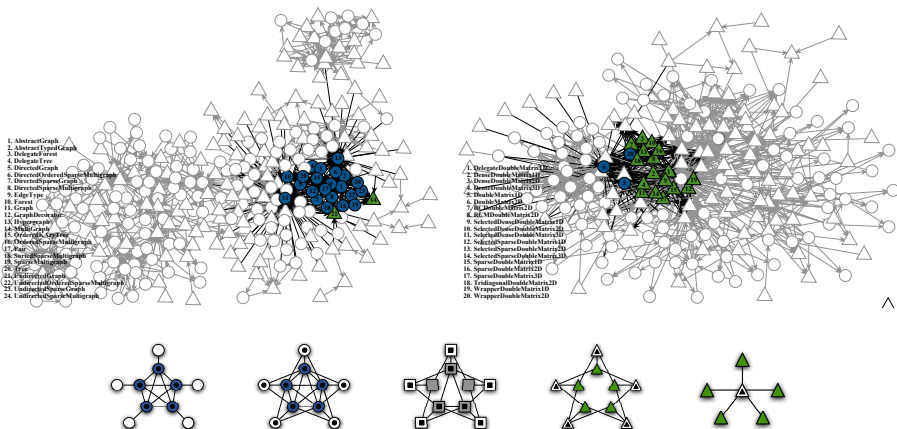


Skupnosti vozlišč le delno sovpadajo s programskim paketi.



Analiza skupin vozlišč

Splošne skupine sovpadajo s paketi ipd. Šubelj and Bajec (2012b); Šubelj et al. (2013b)



^

Uporaba skupin vozlišč

Točnost napovedovanja programskih paketov: Šubelj and Bajec (2012a); Šubelj et al. (2013b)

Omrežje	# Razredov	# Kategorij	Sosedi Γ	Skupine S	Omrežje N	Privzeto	Naključno
<i>jbullet</i>	107	11	72.0%	75.7%	64.5%	28.0%	8.6%
<i>colt</i>	154	16	58.4%	73.4%	55.2%	22.7%	5.9%
<i>jung</i>	237	31	72.2%	74.2%	65.0%	11.4%	3.3%
<i>lucene</i>	1335	178	47.1%	49.2%	43.7%	6.4%	0.6%

Točnost napovedovanja visoko-nivojskih programskih paketov:

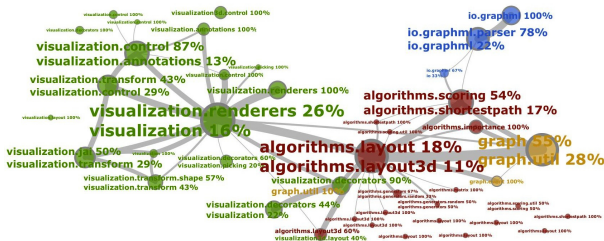
Omrežje	# Razredov	# Kategorij	Sosedi Γ	Skupine S	Omrežje N	Privzeto	Naključno
<i>jbullet</i>	107	5	84.6%	85.0%	78.5%	64.5%	20.4%
<i>colt</i>	154	10	86.4%	83.8%	69.5%	39.0%	9.7%
<i>jung</i>	237	5	89.1%	90.5%	91.1%	44.3%	20.3%
<i>lucene</i>	1335	15	85.5%	90.8%	85.0%	28.2%	6.6%

Točnost napovedovanja vrste, verzije in avtorjev razredov:

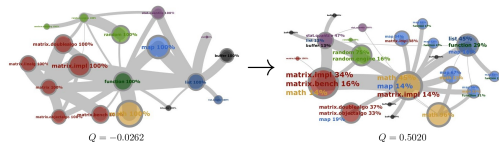
	# Kategorij	Sosedi Γ	Skupine S	Omrežje N	Privzeto	Naključno
Vrsta razreda	2	65.0%	85.2%	84.8%	84.4%	49.9%
Verzija razreda	9	67.7%	72.8%	66.2%	44.3%	11.2%
Avtor razreda	11	71.6%	71.0%	70.9%	44.3%	9.2%

Uporaba skupin vozlišč (II)

Abstrakcija in analiza programskih knjižnic: Šubelj and Bajec (2011d, 2012a)



Reorganizacija programskih paketov (npr. modularno):

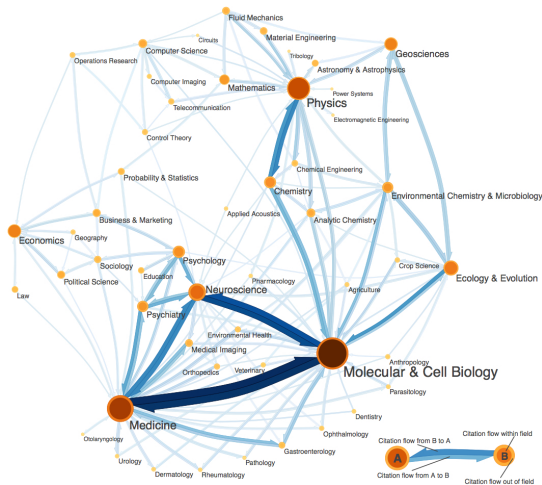


Vsebina

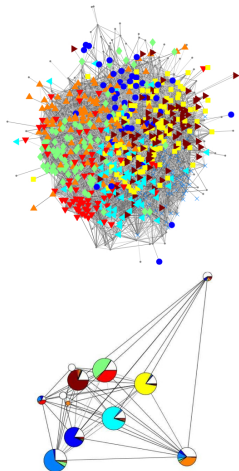
- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Analiza znanstvenih publikacij

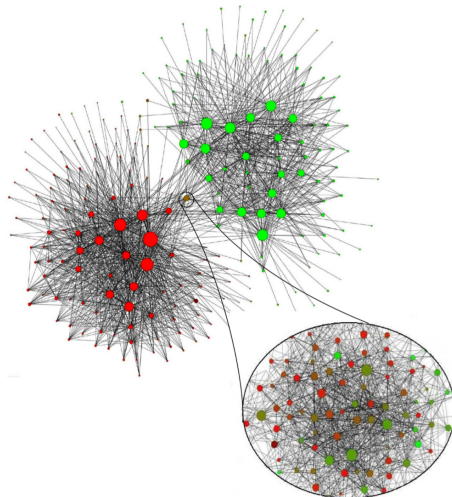
Rosvall and Bergstrom (2008)



Družabna in komunikacijska omrežja



Caltech Porter et al. (2009)



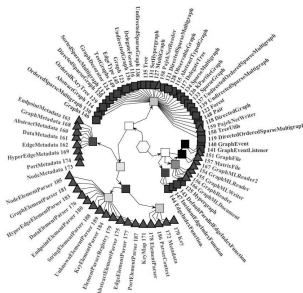
Mobilne komunikacije v Belgiji Blondel et al. (2008)

Vsebina

- 1 Skupine vozlišč v omrežjih
- 2 Odkrivanje skupin vozlišč
 - Osnovna izmenjava oznak
 - Uravnotežena izmenjava oznak
 - Napredna izmenjava oznak
 - Posplošena izmenjava oznak
 - Drugi algoritmi in pristopi
- 3 Uporaba skupin vozlišč
 - Programsko inženirstvo
 - Drugi primeri uporabe
- 4 Zaključek

Zaključek

- pristopi za odkrivanje skupin vozlišč v omrežjih
zahtevnost, splošnost, natančnost, robustnost, enostavnost, brez parametrov
- analiza in uporaba skupin v realnih omrežjih
- za več glej <http://lovro.lpt.fri.uni-lj.si/>



Odkrivanje skupin vozlišč v omrežjih z izmenjavo oznak

Lovro Šubelj

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

<http://lovro.lpt.fri.uni-lj.si/>

lovro.subelj@fri.uni-lj.si

- M. J. Barber and J. W. Clark. Detecting network communities by propagating labels under constraints. *Phys. Rev. E*, 80(2):026129, 2009.
- V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, P10008, 2008.
- M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *P. Natl. Acad. Sci. USA*, 99(12):7821–7826, 2002.
- I. X. Y. Leung, P. Hui, P. Liò, and J. Crowcroft. Towards real-time community detection in large networks. *Phys. Rev. E*, 79(6):066107, 2009.
- M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.
- M. E. J. Newman. The physics of networks. *Phys. Today*, 61(11):33–38, 2008.
- M. E. J. Newman and E. A. Leicht. Mixture models and exploratory analysis in networks. *P. Natl. Acad. Sci. USA*, 104(23):9564, 2007.
- M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Not. Am. Math. Soc.*, 56(9):1082–1097, 2009.
- U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E*, 76(3):036106, 2007.
- M. Rosvall and C. T. Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *P. Natl. Acad. Sci. USA*, 104(18):7327–7331, 2007.
- M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *P. Natl. Acad. Sci. USA*, 105(4):1118–1123, 2008.
- L. Šubelj and M. Bajec. Unfolding network communities by combining defensive and offensive label propagation. In *Proceedings of the ECML PKDD Workshop on the Analysis of Complex Networks*, pages 87–104, Barcelona, Spain, 2010.

- L. Šubelj and M. Bajec. Robust network community detection using balanced propagation. *Eur. Phys. J. B*, 81(3):353–362, 2011a.
- L. Šubelj and M. Bajec. Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction. *Phys. Rev. E*, 83(3):036103, 2011b.
- L. Šubelj and M. Bajec. Generalized network community detection. In *Proceedings of the ECML PKDD Workshop on Finding Patterns of Human Behaviors in Network and Mobility Data*, pages 66–84, Athens, Greece, 2011c.
- L. Šubelj and M. Bajec. Community structure of complex software systems: Analysis and applications. *Physica A*, 390(16):2968–2975, 2011d.
- L. Šubelj and M. Bajec. Software systems through complex networks science: Review, analysis and applications. In *Proceedings of the KDD Workshop on Software Mining*, pages 9–16, Beijing, China, 2012a.
- L. Šubelj and M. Bajec. Ubiquitousness of link-density and link-pattern communities in real-world networks. *Eur. Phys. J. B*, 85(1):32, 2012b.
- L. Šubelj and M. Bajec. Group detection in complex networks: An algorithm and comparison of the state of the art. *Physica A*, 397:144–156, 2014.
- L. Šubelj, N. Blagus, and M. Bajec. Group extraction for real-world networks: The case of communities, modules, and hubs and spokes. In *Proceedings of the International Conference on Network Science*, pages 152–153, Copenhagen, Denmark, 2013a.
- L. Šubelj, S. Žitnik, N. Blagus, and M. Bajec. Node mixing and group structure of complex software networks. *sub. to Adv. Complex Syst.*, page 23, 2013b.
- G. Tibély and J. Kertész. On the equivalence of the label propagation method of community detection and a potts model approach. *Physica A*, 387(19-20):4982–4984, 2008.
- S. Valverde, R. F. Cancho, and R. V. Solé. Scale-free networks from optimal design. *Europhys. Lett.*, 60(4):512–517, 2002.