Network analysis of collaboration in mathematics

Monika Cerinšek, Vladimir Batagelj

Computers in Scientific Discovery 6, Portorož, August 23th 2012

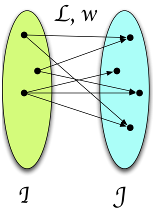
Outline

Two mode networks

Network: $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$; $n = |\mathcal{V}|, m = |\mathcal{L}|$, $w : \mathcal{L} \to \mathbb{R}$ (or some other semiring)

Two-mode network: $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{L}, w)$;
$\mathcal{V} = \mathcal{I} \cup \mathcal{J}$, $(l \in \mathcal{L} \Leftrightarrow l = l(i,j) \wedge i \in \mathcal{I} \wedge j \in \mathcal{J})$

Two-mode network – a network matrix: $\mathbf{W} = [w(i,j)]$ with elements: $w_{i,j} = w(i,j)$ for $(i,j) \in \mathcal{L}$ and $w_{i,j} = 0$ otherwise.



For data from the Web of Science (Knowledge) we can obtain the corresponding networks using the program WoSPajek.
Using other programs we can transform into two-mode networks the data from other bibliographic sources: Scopus, Zentralblatt Math, Google Scholar, IMDB, etc.

Multiplication of networks

Given a pair of compatible networks

$$\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{L}_A, w_A); \mathbf{A}_{\mathcal{I}\times\mathcal{K}}$$

$$\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{L}_B, w_B); \mathbf{B}_{\mathcal{K}\times\mathcal{J}}$$

we call a product of networks $\mathcal{N}_A$ and $\mathcal{N}_B$ a network

$$\mathcal{N}_C = (\mathcal{I}, \mathcal{J}, \mathcal{L}_C, w_C),$$

where $\mathcal{L}_C = \{(i,j); c_{i,j} \neq 0\}$ and $w_C(i,j) = c_{i,j}$ for $(i,j) \in \mathcal{L}_C$.
The product matrix $\mathbf{C} = [c_{i,j}]_{\mathcal{I}\times\mathcal{J}} = \mathbf{A} * \mathbf{B}$ is defined in the standard way

$$c_{i,j} = \sum_{k\in\mathcal{K}} a_{i,k} \cdot b_{k,j}$$

In the case when $\mathcal{I} = \mathcal{K} = \mathcal{J}$ we are dealing with ordinary one-mode networks (with square matrices).

Normalized co-authorship network

Let **WA** be the works $\times$ authors two mode network with weights $wa_{pi} \in \{0, 1\}$ for work p and author i.

$$\Rightarrow \forall p \in W : \sum_{i \in A} wa_{pi} = \mathrm{outdeg}(p)$$

Normalized version **N** with weights $n_{pi} = wa_{pi}/\mathrm{outdeg}(p)$.

$$\Rightarrow \forall p \in W : \sum_{i \in A} n_{pi} = 1$$

Transformations of networks

Let $\mathcal{N}$ be a network.

Binarization $b(\mathcal{N})$
All weights are set to 1.

Transposition $t(\mathcal{N})$
Direction of all arcs is reversed.

Row-normalization $n(\mathcal{N})$
The weight of each arc a is divided by the sum of weights of all arcs having the same initial vertex as the arc a. For binary networks

$$n(\mathbf{A}) = \mathrm{diag}\left(\frac{1}{outdeg(i)}\right)_{i\in\mathcal{I}} * \mathbf{A}$$

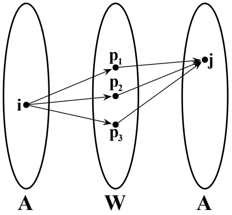
...

From **N** we can get **WA** as $\mathbf{WA} = b(\mathbf{N})$.

Ordinary collaboration network

$$\mathbf{Co} = \mathbf{AW} * \mathbf{WA}$$

$$co_{ij} = \sum_{p \in W} aw_{ip} wa_{pj} = \sum_{p \in N(i) \cap N(j)} 1$$

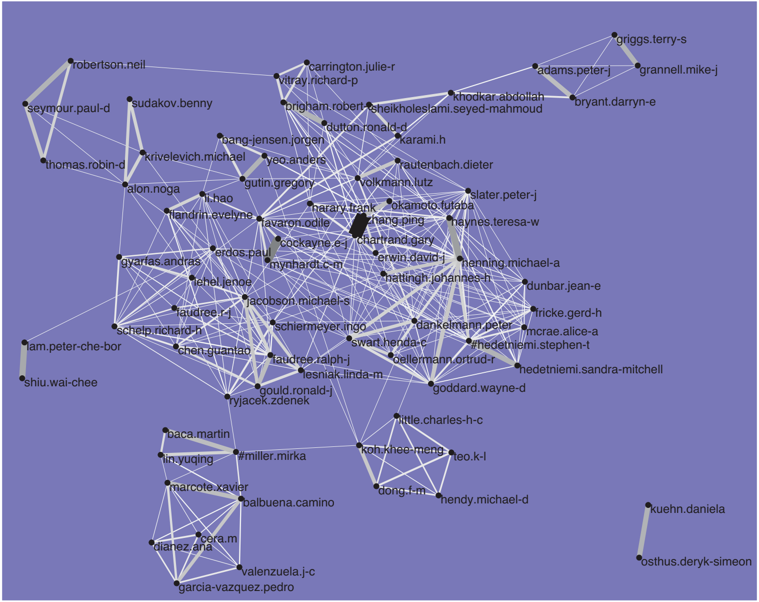


co_{ij} = the number of works that authors i and j wrote together
It holds: $co_{ij} = co_{ji}$.

Problem: The **Co** network is composed of complete graphs on the set of work's authors. Works with many authors produce large complete subgraphs.

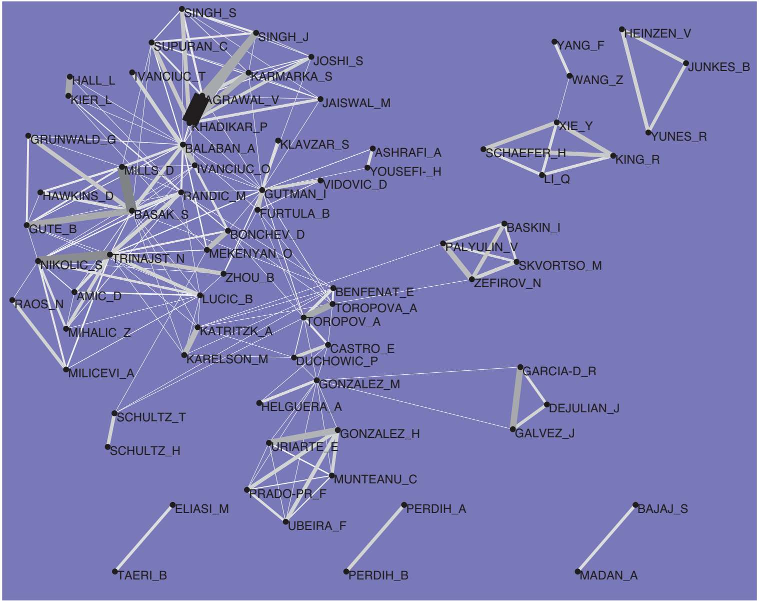
Valued cores in collaboration network – Graph theory

$t = 34$



Valued cores in collaboration network – Topological indices

$t = 54$



Second collaboration network

$\mathbf{Cn} = t(\mathbf{WA}) * \mathbf{N}$

$$cn_{ij} = \sum_{p \in W} wa_{pi} n_{pj} = \sum_{p \in N(i) \cap N(j)} n_{pj}$$

cn_{ij} = contribution of author j to works, that (s)he wrote together with the author i.

$cn_{ii} = \sum_{p \in N(i)} n_{pi}$ is the contribution of author i to all of his/her works.

It holds

$$\sum_{i \in A} \sum_{j \in A} wa_{pi} n_{pj} = \mathrm{outdeg}(p), \quad \sum_{j \in A} cn_{ij} = \mathrm{indeg}(i).$$

If $n_{pi} = \frac{wa_{pi}}{\mathrm{outdeg}(p)}$, then $cn_{ij} = cn_{ji}$.

Second collaboration network

self-sufficiency: $S_i = \dfrac{cn_{ii}}{\mathrm{indeg}(i)}$

collaborativness (co-authorship index): $K_i = 1 - S_i$

It holds:

$$\sum_{i \in A} \sum_{j \in A} cn_{ij} = \sum_{i \in A} \mathrm{indeg}(i) = 2m_{Co}$$

The "best" authors in Graph theory

Author	Contribution of the author	Contribution of other co-authors	Collaborativness	No. of all articles
Volkmann, Lutz	123.55	92.45	42.80%	216
Henning, Michael A.	110.87	121.13	52.21%	232
Liu, Yanpei	102.42	93.58	47.75%	196
Alon, Noga	85.39	91.61	51.76%	177
Tuza, Zsolt	77.05	72.95	48.63%	150
Zhu, Xuding	76.85	55.15	41.78%	132
Gutman, Ivan	70.13	72.87	50.96%	143
Thomassen, Carsten	68.83	13.17	16.06%	82
Mohar, Bojan	67.85	43.15	38.87%	111
Liu, Guizhen	67.28	69.72	50.89%	137
Liu, Bolian	63.67	55.33	46.50%	119
Klavžar, Sandi	62.94	67.06	51.58%	130
Bollobás, Béla	62.53	80.47	56.27%	143
Kostochka, Alexandr V.	62.18	77.82	55.58%	140
Zhang, Ping	60.47	96.53	61.49%	157
Li, Xueliang	60.40	75.60	55.59%	136
Rödl, Vojtěch	57.98	88.02	60.29%	146
Zhang, Zhongfu	57.07	104.93	64.77%	162
Zelinka, Bohdan	56.50	2.50	4.23%	59
McKee, Terry A.	54.98	9.02	14.09%	64

The "best" authors in Topological indices

Author	Contribution of the author	Contribution of other co-authors	Collaborativness	No. of all articles
Randić, Milan	402.69	159.31	28.35%	562
Gutman, Ivan	363.64	269.36	42.55%	633
Balaban, Alexandru	211.83	125.17	37.14%	337
Ivanciuc, Ovidiu	168.19	63.81	27.50%	232
Diudea, Mircea	157.23	101.77	39.30%	259
Estrada, Ernesto	142.86	68.14	32.29%	211
Basak, Subhash	139.17	157.83	53.14%	297
Pyka, Andreas	121.40	36.60	23.17%	158
Trinajstić, Nenad	112.66	198.35	63.78%	311
Chou, Kuo-Chen	112.05	94.95	45.87%	207
Kier, Lemont	107.81	59.19	35.44%	167
Ashrafi, Ali Reza	103.98	114.02	52.30%	218
Bonchev, Danail	102.85	82.15	44.41%	185
Khadikar, Padmakar	100.80	207.20	67.27%	308
Hosoya, Haruo	94.49	56.51	37.42%	151
Torrens, Francisco	93.17	60.83	39.50%	154
Zhou, Bo	84.21	60.79	41.92%	145
Pogliani, Lionello	82.48	4.52	5.20%	87
Toropov, Andrey	79.32	102.68	56.42%	182
Lukovits, István	72.05	33.95	32.03%	106

Third collaboration network

$\mathbf{Ct} = t(\mathbf{N}) * \mathbf{N}$

ct_{ij} = the total contribution of collaboration of authors i and j to works.
It holds

$$ct_{ij} = ct_{ji}, \quad \sum_{i \in A} \sum_{j \in A} t(n)_{ip} n_{pj} = 1.$$

The total contribution of author i to works from W is

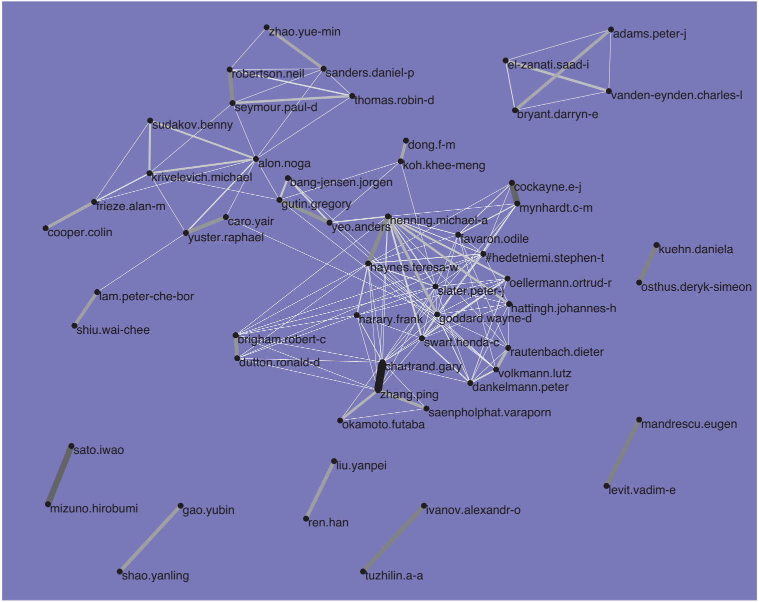
$$\sum_{j \in A} ct_{ij} = \sum_{p \in W} n_{pi}.$$

It holds

$$\sum_{i \in A} \sum_{j \in A} ct_{ij} = |W|.$$

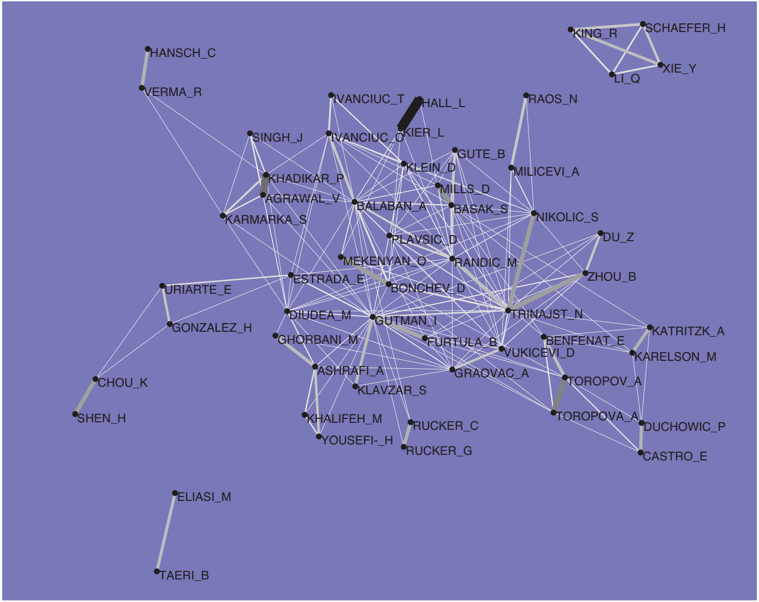
Valued cores for Ct – Graph theory

$t = 4$



Valued cores for Ct – Topological indices

$t = 3$



Thank you for your attention.

Vladimir Batagelj, vladimir.batagelj@fmf.uni-lj.si
Monika Cerinšek, monika@hruska.si