Analysis of enriched scientific collaboration networks

Miloš Savić and Mirjana Ivanović

Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Serbia

email: {svc,mira}@dmi.uns.ac.rs

Introduction

- *Scientific collaboration is a social process and probably there are as many reasons for researchers to collaborate as there are reasons for people to communicate* (Katz)
- Co-authorship – the most visible and well-documented manifestation of scientific collaboration
- Co-authorship networks – social networks encompassing researchers
  - ◊ Nodes – researchers
  - ◊ A and B are connected if A and B co-authored at least one publication (with or without other co-authors)
  - ◊ Link weights – the strength of research collaboration

Enriched co-authorship networks are co-authorship networks in which nodes are annotated with

1. attributes indicating demographic characteristics of researchers (e.g. gender, age, academic position, and so on), and
2. researcher evaluation metrics that quantify various determinants of research performance.

Methodology

Our methodology to study the structure and evolution of enriched co-authorship networks includes and combines:

1. domain-independent metrics and methods used in complex network analysis,
2. domain-dependent researcher evaluation metrics,
3. non-parametric statistical tests applied to the sets of metric values of independent groups of nodes in an enriched co-authorship network, and
4. graph representations derived from enriched co-authorship networks.

Case study

- The intra-institutional research collaboration network of the Faculty of Sciences, University of Novi Sad (FS-UNS)
- The network was extracted from the bibliographic records stored in the institutional CRIS system
- 423 researchers from 5 departments (DBE, DP, DG, DC and DMI)
- Discrete node attributes: gender, departmental affiliation
- Numeric node attributes
  - research productivity measured by the normal, fractional and straight counting method
  - SRCI – Serbian Research Competency Index
  - collaboration metrics quantifying local and external research collaboration
  - node centrality metrics quantifying institutional importance

Domain-independent metrics and methods

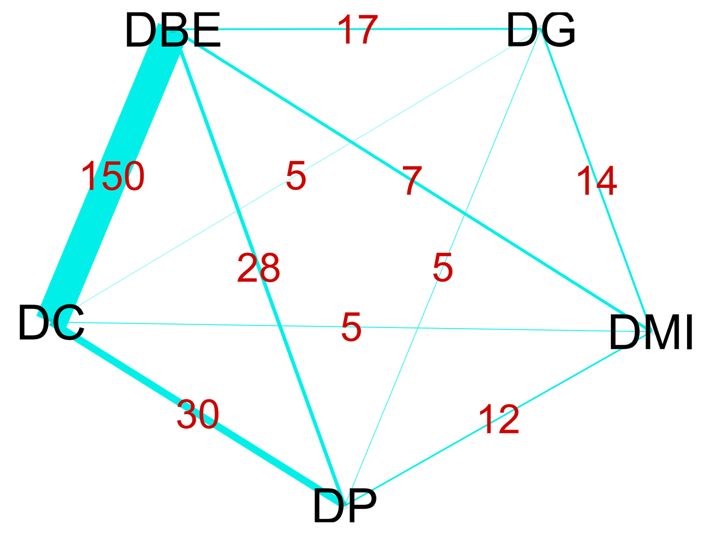
- Centrality metrics and algorithms (ranking researchers and research groups by their position in collaboration networks)
- Network decomposition methods – k-core decomposition (core and periphery researchers), community detection techniques (identification of research groups)
- Graph clustering evaluation (GCE) metrics – cohesiveness of research communities and categorically induced subgraphs

Block models of enriched co-authorship networks

- P – a partition of the nodes into k groups
- Block model corresponding to P is a network of node groups
- Block models derived according to categorical node attributes (e.g. departmental affiliation)
- Block models derived according to partitions obtained after community detection

FS-UNS block models

The departmental collaboration network of FS-UNS is a clique with highly unbalanced link weights



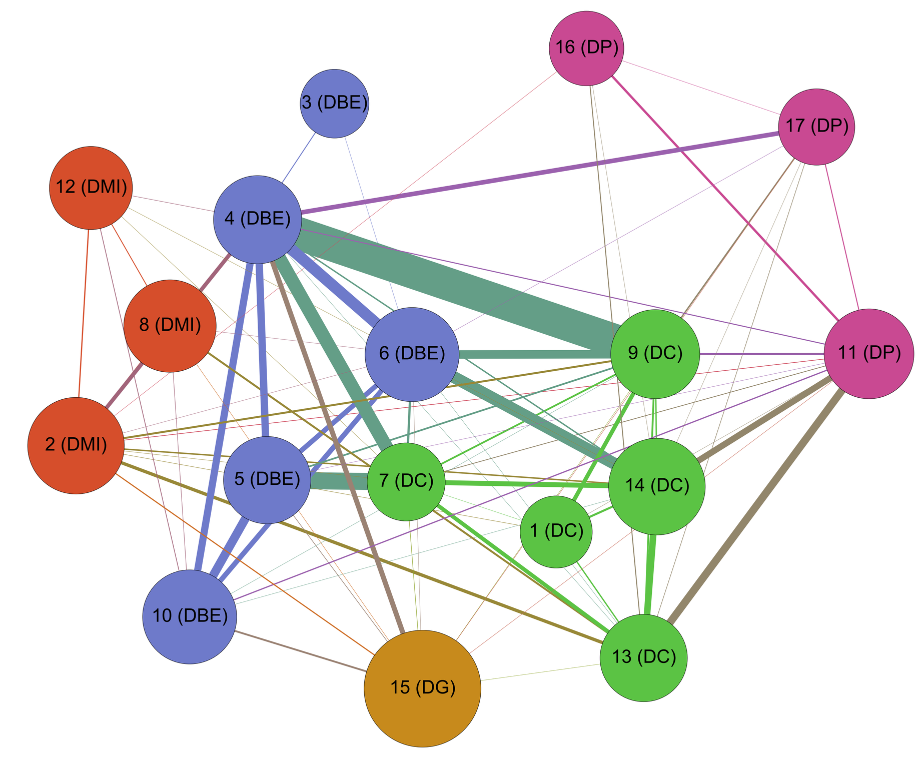
GCE metrics: FS-UNS departments are cohesive clusters in the FS-UNS network

Metric	DBE	DP	DC	DMI	DG
Intra-dept links	660	240	617	197	560
Inter-dept links	412	174	411	71	96
W(intra-dept links)	8073	5636	9261	1532	2513
W(inter-dept links)	1607	683	1825	195	160
Conductance	0.17	0.11	0.16	0.11	0.06
Flake degree fraction	0.97	0.93	0.98	0.95	0.95

Identification of research groups

Six community detection techniques, the best partitioning is obtained by the Louvain algorithm

Algorithm	Q	NC	w^{intra}	w^{inter}	r
GMO	0.8371	18	6919.45	655.66	0.0947
IM	0.8141	41	6618.53	956.58	0.1445
LV	0.8466	17	6920.37	654.74	0.0946
WT	0.8207	37	6873.07	702.04	0.1021
EB	0.5486	13	5248.49	2326.63	0.4433
SOM	0.6022	27	6466.84	1108.28	0.1714



Metric-based comparison test

- X and Y – two independent subset of nodes in a enriched co-authorship network
- T – a probability threshold indicating a strong stochastic dominance
- **metric-based-comparison-test**(X, Y, T)**:**

  for-each numeric node attribute M:

  $M(X)$ – the set of M values for nodes in X

  $M(Y)$ – the set of M values for nodes in Y

  p = apply the MWU test to $M(X)$ and $M(Y)$

  if the null hypothesis rejected ($p < 0.05$):

  compute PS_X and PS_Y

  $PS_X = P(x > y)$, $PS_Y = P(y > x)$

  x – a randomly selected value from X

  y – a randomly selected value from Y

  if $PS_X > T$ or $PS_Y > T$:

  report not only statistically significant differences between X and Y regarding M, but also a strong tendency of superiority

Group superiority graphs (GSG) are directed graphs reflecting stochastic dominance among node groups with respect to numerical node attributes.

- M – a numeric node attribute, A and B two node groups
- $A \to B$ in the GSG of M if nodes in A strongly tend to have higher values of M than nodes in B
- **GSGs are derived from block models according to the metric-based comparison test, one GSG per research evaluation metrics attached to nodes**

Main results

The application of the metric-based comparison test to the FS-UNS enriched co-authorship network revealed that:

- FS-UNS researchers involved in inter-department collaborations (G_1) strongly tend to be more productive, collaborative and institutionally important compared to FS-UNS researchers whose collaboration is bounded to their own research departments (G_2).

Metric	Avg(G_1)	Avg(G_2)	U	p	PS$_1$	PS$_2$
PRON	104.9031	32.9031	10333	$< 10^{-4}$	0.76	0.23
PROS	29.2555	13	13781	$< 10^{-4}$	0.68	0.29
PROF	27.9682	12.3087	13477.5	$< 10^{-4}$	0.7	0.3
SRCI	160.378	58.6939	11178.5	$< 10^{-4}$	0.75	0.25
COLL	69.7313	20.9337	7360	$< 10^{-4}$	0.83	0.16
LCOLL	18.7225	7.4592	7486.5	$< 10^{-4}$	0.82	0.16
ECOLL	51.0088	13.4745	8411.5	$< 10^{-4}$	0.8	0.18
BET	769.6687	98.0929	7775	$< 10^{-4}$	0.82	0.17

- Core FS-UNS researchers strongly tend to be more productive, collaborative and institutionally important compared to FS-UNS researchers located on the periphery of the network.

Metric	Avg(C)	Avg(P)	U	p	PS$_1$	PS$_2$
PRON	124.2576	49.6354	7954	$< 10^{-4}$	0.78	0.22
PROF	31.7894	16.1697	10003	$< 10^{-4}$	0.73	0.27
PROS	32.4924	17.3430	10647	$< 10^{-4}$	0.70	0.28
SRCI	172.8083	89.7819	9684	$< 10^{-4}$	0.74	0.26
COLL	88.2273	29.8051	4653	$< 10^{-4}$	0.87	0.13
LCOLL	26.4697	8.0072	784.5	$< 10^{-4}$	0.98	0.02
ECOLL	61.7576	21.7978	7191	$< 10^{-4}$	0.80	0.19
BET	813.9461	312.2748	8040	$< 10^{-4}$	0.78	0.22

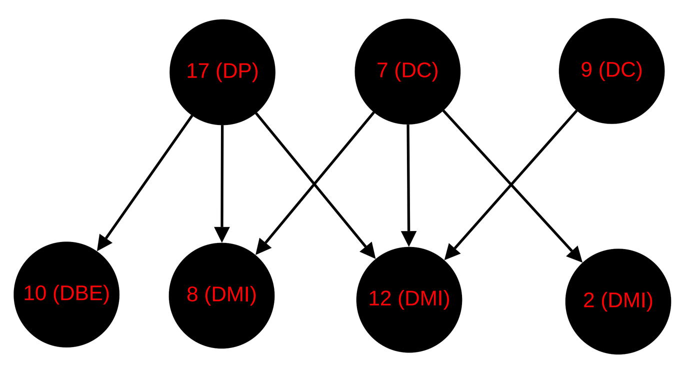
- FS-UNS researchers involved in inter-group collaborations (G_1) strongly tend to be more productive collaborative and institutionally important compared to FS-UNS researchers who collaborate only with colleagues from their own research groups (G_2)

Metric	Avg(G_1)	Avg(G_2)	U	p	PS$_1$	PS$_2$
PRON	98.6367	26.8662	8221.5	$< 10^{-4}$	0.78	0.21
PROF	26.5373	11.1954	11043	$< 10^{-4}$	0.71	0.29
PROS	27.8801	11.6127	11325	$< 10^{-4}$	0.69	0.28
SRCI	151.3667	51.1648	9037.5	$< 10^{-4}$	0.76	0.24
COLL	65.1873	17.5845	5301	$< 10^{-4}$	0.86	0.14
LCOLL	17.4569	7.4014	6744.5	$< 10^{-4}$	0.81	0.16
ECOLL	47.7303	10.1831	5932	$< 10^{-4}$	0.84	0.15
BET	687.4335	73.2130	5683	$< 10^{-4}$	0.84	0.14

FS-UNS group superiority graphs

Metric	Nodes	Links	Superior groups	Inferior groups	Bipartite
PRON	7	7	3	4	yes
PROF	0	0	/	/	/
PROS	0	0	/	/	/
SRCI	11	10	2	9	yes

The group superiority graph corresponding to the productivity measured by normal publication counting (PRON):



Conclusions

- Our case study demonstrates that the proposed methodology enables an in-depth analysis of research collaboration and its relationships with other indicators of research performance
- Researchers located in the core of a co-authorship network and researchers involved in inter-department and inter-group collaborations tend to be highly productive
- Our case study indicates that PRON and SRCI are biased measures of research productivity

References

1. M. Savić, M. Ivanović, L. C. Jain. *Complex Networks in Software, Knowledge, and Social Systems*. Springer International Publishing, 2019.
2. M. Savić, M. Ivanović, B. Dimić Surla. Analysis of intra-institutional research collaboration: a case of a Serbian faculty of sciences. *Scientometrics* 110(1): 195–216, 2017.
3. M. Savić, M. Ivanović, M. Radovanović, B. Dimić Surla. Gender-Based Analysis of Intra-Institutional Research Productivity and Collaboration. *Fundamenta Informaticae* 162(4):237–258, 2018.