Temporal Network Assortativity

Moses A. Boudourides[1]

Visiting Professor of Mathematics, New York University Abu Dhabi

[1] Moses.Boudourides@nyu.edu

with Giannis Tsakonas & Sergios Lenis

جامعة نيويورك أبوظبي

NYU ABU DHABI

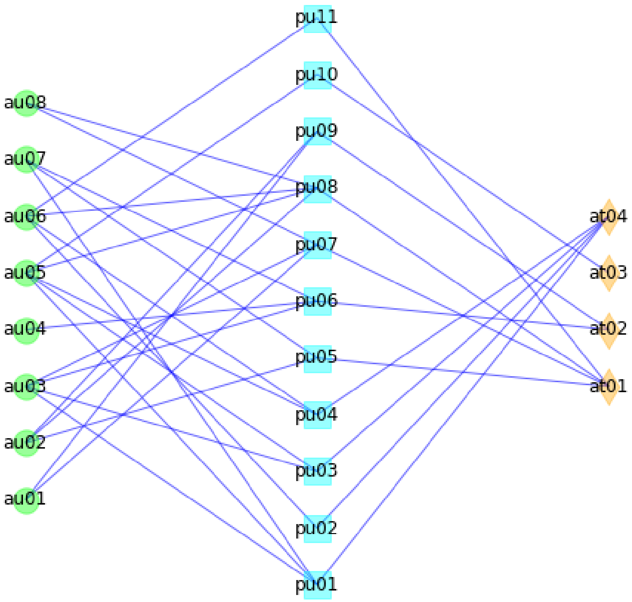
NYUAD RGNS, October 13, 2019

The Method

Let us start with the stationary case of a graph $G = (V, E)$, in which the set of nodes V is composed of three different (disjoint) sets, $V = V_1 \cup V_2 \cup V_3$, where:

- V_1 is a set of n_1 **authors**,
- V_2 is a set of n_2 **publications**,
- V_3 is a set of n_3 **access types** of the sources, in which publications are classified (tpically, $n_3 = 4$).

Partial Tripartite Graph of Authors-Publications-Access Types



Now, if the only existing edges in the set E are between authors–publications and publications–access types (but there are no edges either among authors or among publications or among access types), G is a **partial tripartite graph** of authors–publications–access types. Denoting by **B** the **biadjacency matrix** of authors–publications and by **C** the **biadjacency matrix** of publications–access types, the **adjacency matrix** of the tri-partite graph G is the following:

$$\begin{array}{r c}
 & \begin{array}{ccc} \text{Authors} & \text{Publications} & \text{Access Types} \end{array} \\
\begin{array}{r} \text{Authors} \\ \text{Publications} \\ \text{Access Types} \end{array} & \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \hline \mathbf{B}^T & \mathbf{0} & \mathbf{C} \\ \hline \mathbf{0} & \mathbf{C} & \mathbf{0} \end{array}\right]
\end{array}$$

Assuming each publication is of a unique access type, the the biadjacent matrix **C** is *one–to–one*, i.e.,

$$\sum_{j \in V_3} C_{ij} = 1, \text{ for all } i \in V_2,$$

and, without any loss of generality, the adjacency matrix of the partial tripartite graph of authors–publications–access types is written as:

$$\begin{array}{cc}
 & \begin{array}{ccc} \text{Authors} & \text{Publications} & \text{Access Types} \end{array} \\
\begin{array}{c} \text{Authors} \\ \text{Publications} \\ \text{Access Types} \end{array} & \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{B} & \mathbf{BC} \\ \hline \mathbf{B}^T & \mathbf{0} & \mathbf{C} \\ \hline \mathbf{C}^T\mathbf{B}^T & \mathbf{C} & \mathbf{0} \end{array}\right]
\end{array}$$

At this point, we are going to reduce the partial tripartite graph of authors–publications–access types to a a bipartite graph of authors–publications, in which the access type is encoded as an attribute of authors. To do so, we need to modify the categories (values) of the access type attribute.

Actually, as an author may have published multiple publications, each publication appearing in sources classified by various access types, what should be associated as an attribute to this author is not *a single one* of the n_3 values of the set V_3 of access type attributes but, in general, a single *subset* among the 2^{n_3} subsets of all possible combinations of the n_3 access types. In particular, to find the assignment of a set of access types to an author, we need to proceed as follows:

Let $H_{13} = (V_1 \cup V_3, E_{13})$ be the bipartite graph of authors–access types, the biadjacency matrix of which is **BC**.

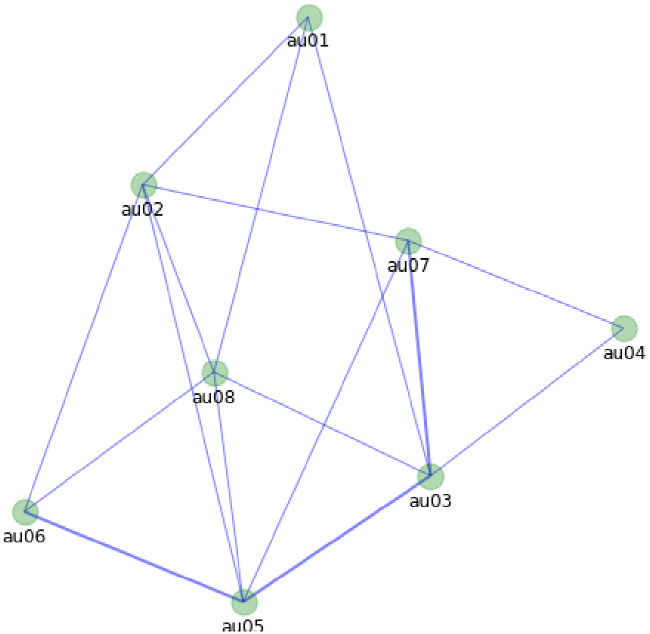
For any $v \in V_1$, let $\alpha(v)$ be a subset of the power set 2^{V_3} of all subsets of elements of V_3 such that, for any $a \in \alpha(v), (v, a)$ in an edge in H_{13}, i.e., $(a, v) \in E_{13}$.

Then, the access type on authors is obtained through the set–valued mapping $\alpha \colon V_1 \to 2^{V_3}$.

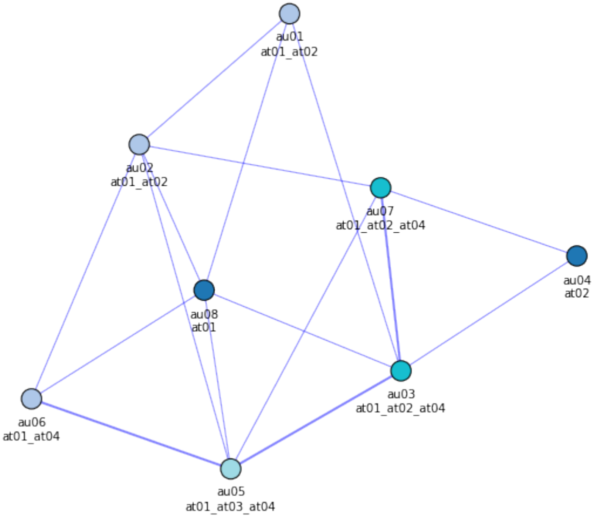
In particular, this means that in the bipartite graph of authors–publications $H_{12} = (V_1 \cup V_2, E_{12})$, the biadjacency matrix of which is $\mathbf{B}$, the mode of authors V_1 is attributed by α.

Therefore, projecting H_{12} on the mode of authors generates the one-mode graph G_α, commonly called **co-authorship network**, the adjacency matrix of which is $\mathbf{B}\mathbf{B}^T$, possessing the attribute α of set–valued (combinations of) access types on all its nodes (authors).

The Co-Authorship Graph



The Co-Authorship Graph under 6 Combined Access Types
Access Type Assortativity Coefficient = -0.098

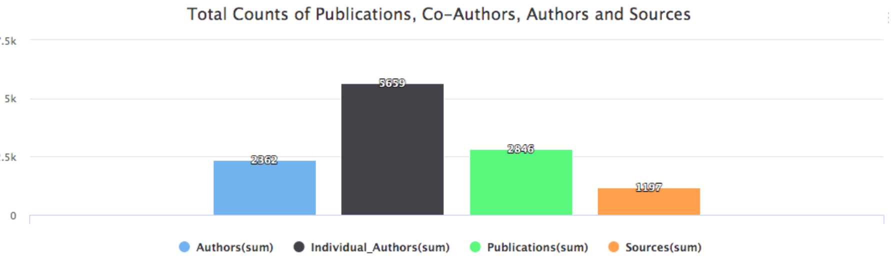


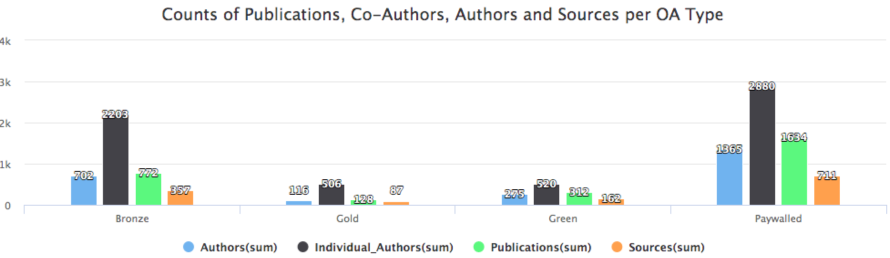
The Dataset

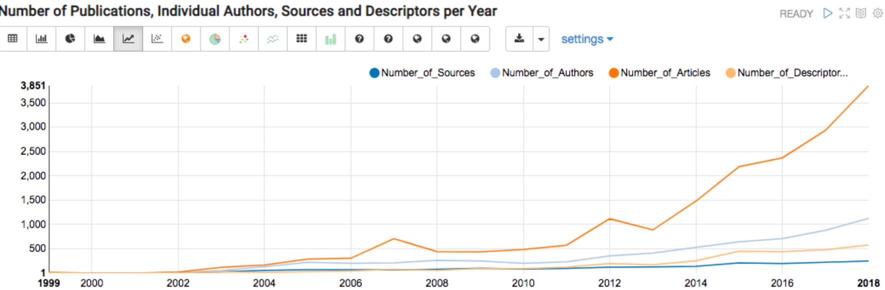
Input

- 2,846 publications / March 2019
- Data retrieved from / structured data / OA versioning scheme
- Timespan: 1999-2018

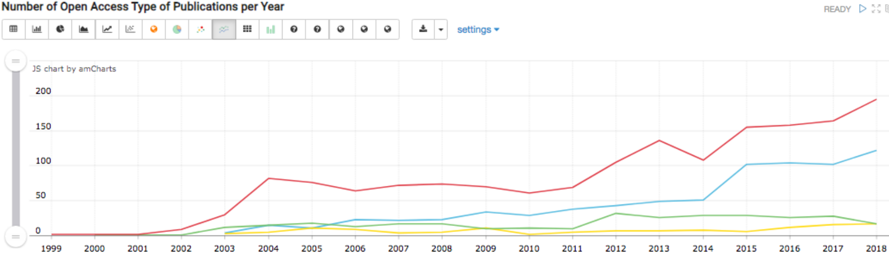
- Basic query: `TITLE: ("Open Science") OR TITLE: ("Open Access")`
- Refinements: `NOT TOPIC: (endoscop*) NOT TOPIC: (fish*) NOT TOPIC: (enteroscop*) NOT TOPIC: (schedul*) ...`


Total Counts of Publications, Co-Authors, Authors and Sources
.5k
5k
.5k
0
5659
2362
2846
1197
Authors(sum)
Individual_Authors(sum)
Publications(sum)
Sources(sum)

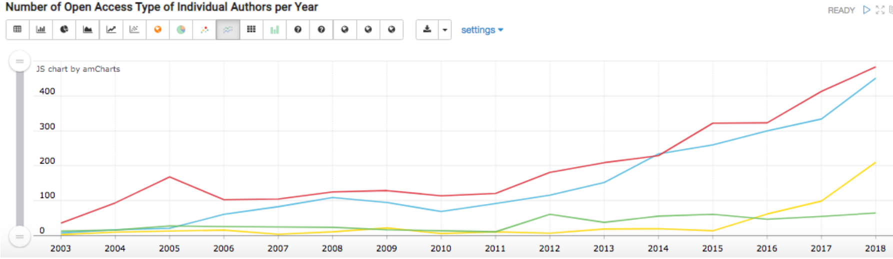

Counts of Publications, Co-Authors, Authors and Sources per OA Type
4k
3k
2k
1k
0
702
2203
772
357
116
506
128
87
275
520
312
162
1365
2880
1634
711
Bronze
Gold
Green
Paywalled
Authors(sum)
Individual_Authors(sum)
Publications(sum)
Sources(sum)


Number of Publications, Individual Authors, Sources and Descriptors per Year
READY
settings
Number_of_Sources
Number_of_Authors
Number_of_Articles
Number_of_Descriptor...
3,851
3,500
3,000
2,500
2,000
1,500
1,000
500
1
1999
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018

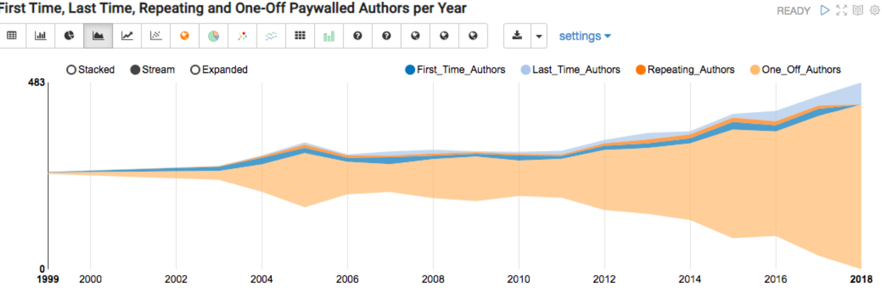
Number of Open Access Type of Publications per Year



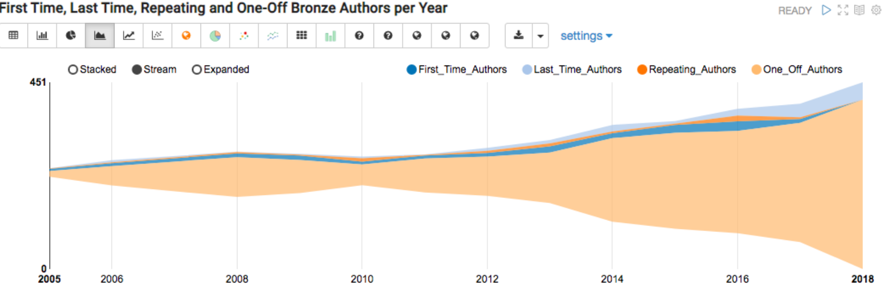
Number of Open Access Type of Individual Authors per Year



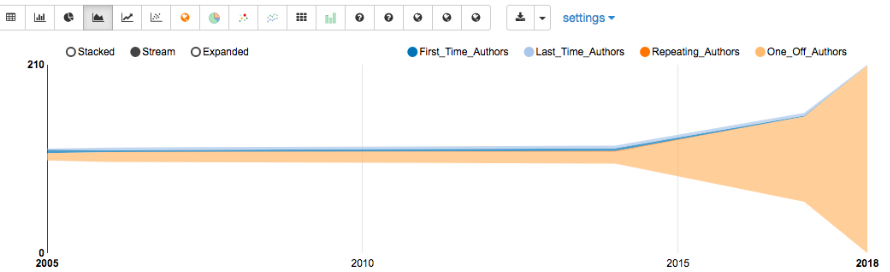
First Time, Last Time, Repeating and One-Off Paywalled Authors per Year



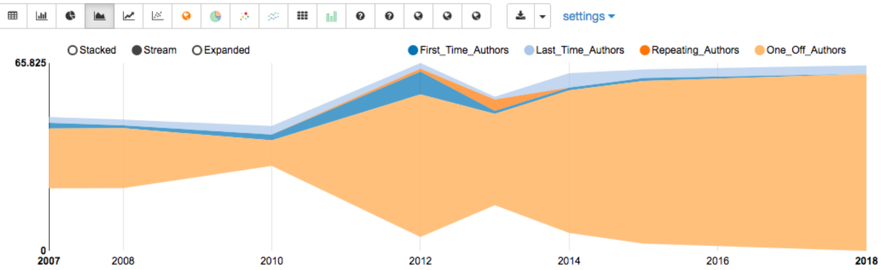
First Time, Last Time, Repeating and One-Off Bronze Authors per Year



First Time, Last Time, Repeating and One-Off Gold Authors per Year

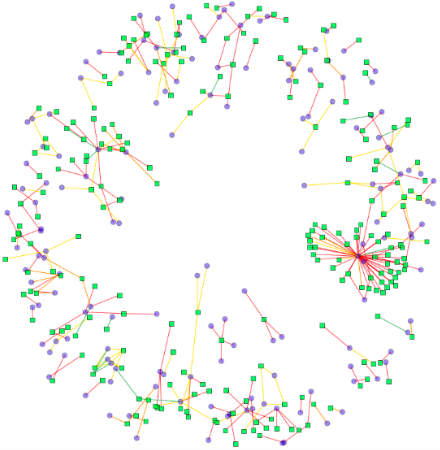


First Time, Last Time, Repeating and One-Off Green Authors per Year

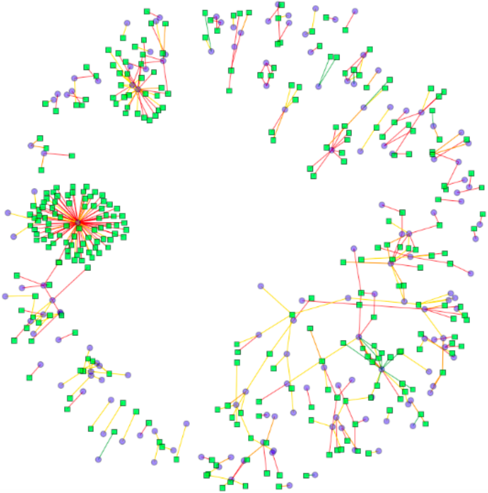


The Bipartite Graph of Authors-Publications for 155 Authors publishing before and after 2011

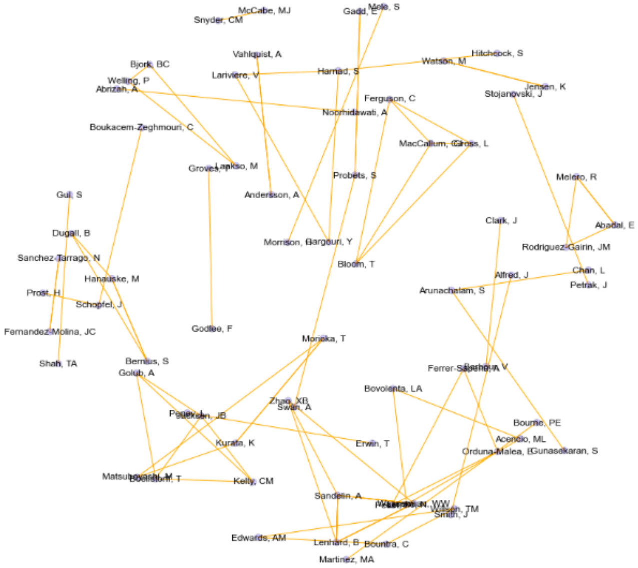
The bipartite graph of authors-publications in Open Science from 1999 to 2011
(blue circles = 155 authors, lime squares = 231 publications,
red links = 159 paywalled sources, bronze links = 32 bronze sources,
gold links = 80 gold sources, green links = 17 green sources, 110 connected components)



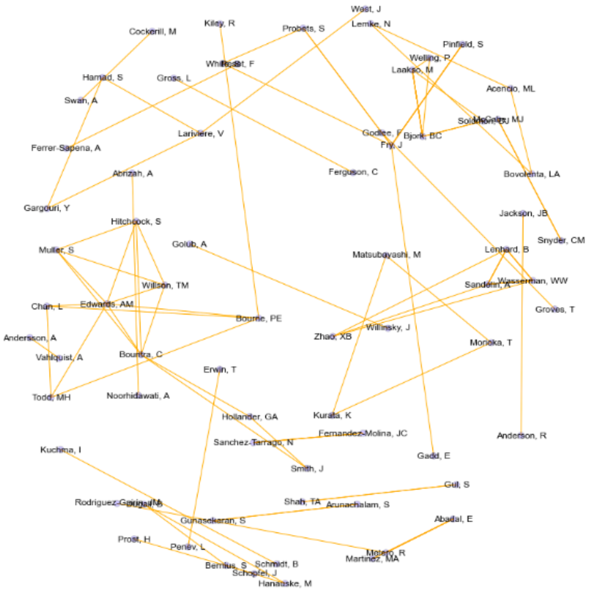
The bipartite graph of authors-publications in Open Science from 2012 to 2018
(blue circles = 155 authors, lime squares = 314 publications,
red links = 209 paywalled sources, bronze links = 34 bronze sources,
gold links = 129 gold sources, green links = 18 green sources, 108 connected components)



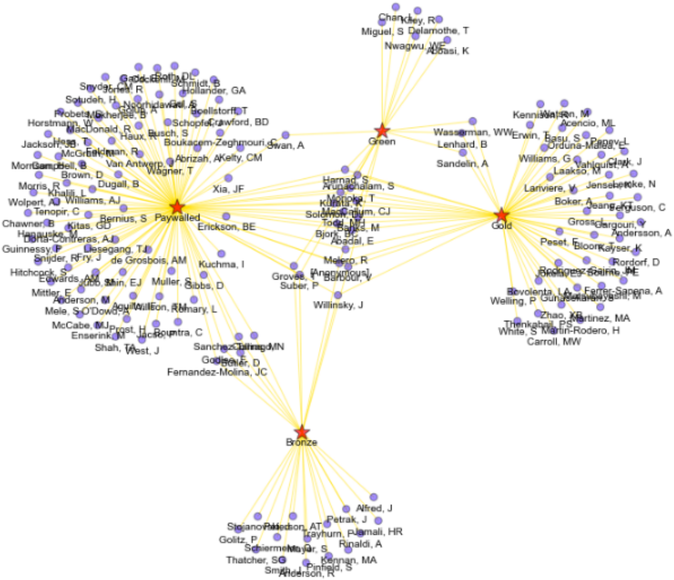
The coauthorship graph of publications in Open Science from 1999 to 2018
(72 authors, 76 coauthorships as orange links, 27 connected components)



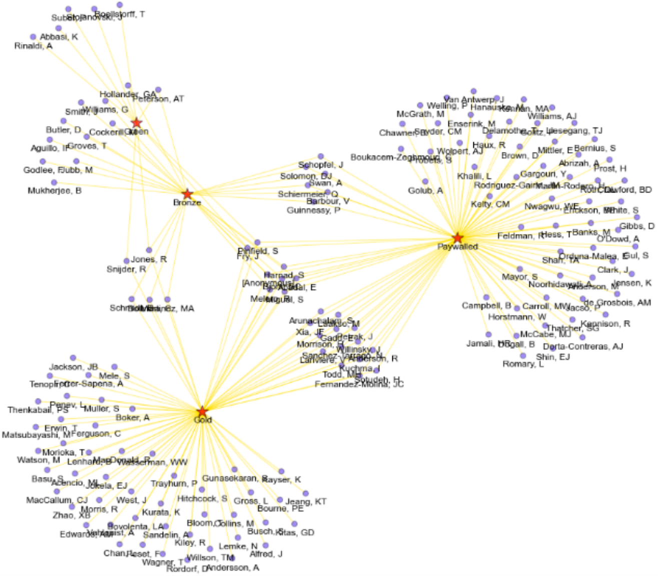
The coauthorship graph of publications in Open Science from 2012 to 2018 (71 authors, 99 coauthorships as orange links, 24 connected components)



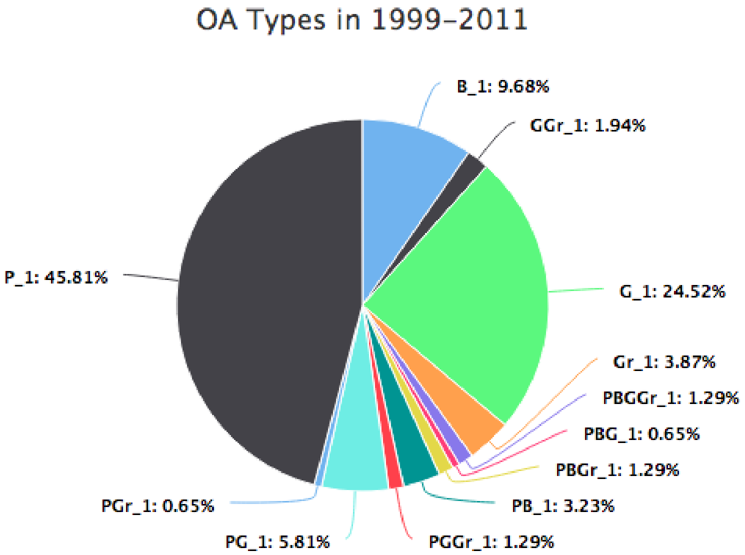
The bipartite graph of authors-OA_types in Open Science from 1999 to 2011
(blue circles = 155 authors, red stars = 4 Open Access types,
gold links = 189 authors-OA_types associations)



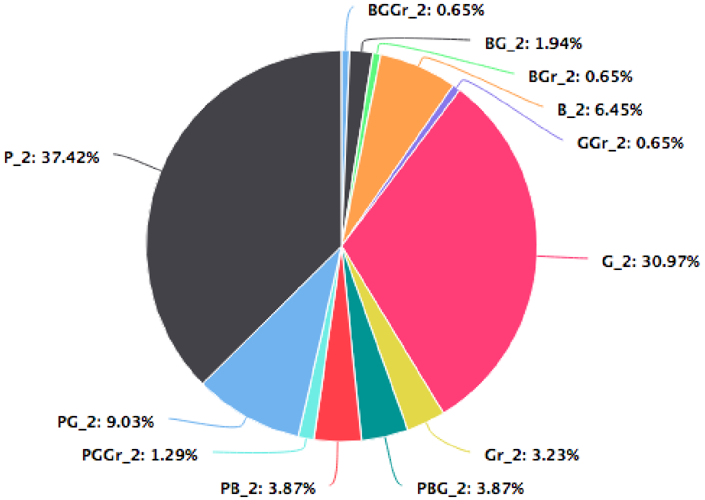
The bipartite graph of authors-OA_types in Open Science from 2012 to 2018
(blue circles = 155 authors, red stars = 4 Open Access types,
gold links = 198 authors-OA_types associations)



Mixed OA Types



OA Types in 2012–2018



Assortativity of a Partition

- Let $\mathcal{P} = \{P_1, P_2, \ldots, P_p\}$ be a vertex partition of graph G.
- Identifying $\mathcal{P}$ to a p–assignment $\mathcal{A}_\mathcal{P}$ of enumerative attributes to the vertices of G , one can define (cf. **Mark Newman**, 2003), the (normalized) **enumerative attribute assortativity** (or **discrete assortativity**) **coefficient of partition** $\mathcal{P}$ as follows:

$$r_\mathcal{P} = r_\mathcal{P}(\mathcal{A}_\mathcal{P}) = \frac{\operatorname{tr} \mathbf{M}_\mathcal{P} - ||\mathbf{M}^2_\mathcal{P}||}{1 - ||\mathbf{M}^2_\mathcal{P}||},$$

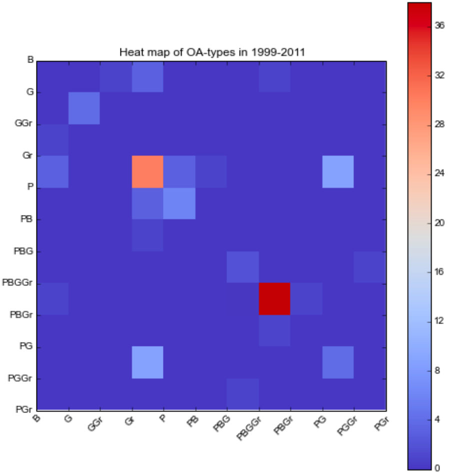
where $\mathbf{M}_\mathcal{P}$ is the $p \times p$ (normalized) **mixing matrix of partition** $\mathcal{P}$. Equivalently:

$$r_\mathcal{P} = \frac{\sum_{i,j \in V}(A_{ij} - \frac{k_i k_j}{2m})\delta(\mathcal{A}_\mathcal{P}(i), \mathcal{A}_\mathcal{P}(j))}{2m - \sum_{i,j \in V}(\frac{k_i k_j}{2m})\delta(\mathcal{A}_\mathcal{P}(i), \mathcal{A}_\mathcal{P}(j))},$$

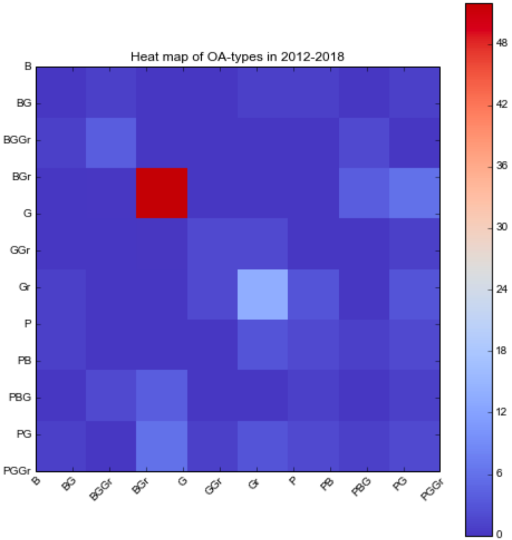
where $\{A_{ij}\}$ is the adjacency matrix of graph G, m is the total number of edges of G, k_i is the degree of vertex i and $\delta(x, y)$ is the Kronecker delta.

Mixing and OA Type Assosartivity Coefficient

The OA-type assortativity coefficient of the co-authorship network in 1999-2011 is $= 0.563$



The OA-type assortativity coefficient of the co-authorship network in 2012-2018 is $= 0.406$



Authors' Swings

The OA-type assortativity coefficient of the graph of authors' swings before and after 2011 is $= 0.001$ *(non-assortative graph)*