

Group Identification in Online Communities

Balázs Sziklai

Workshop on Future Directions in Computational Social Choice, Budapest

e-mail: sziklai.balazs@rtk.mta.hu

2016. 11. 21.



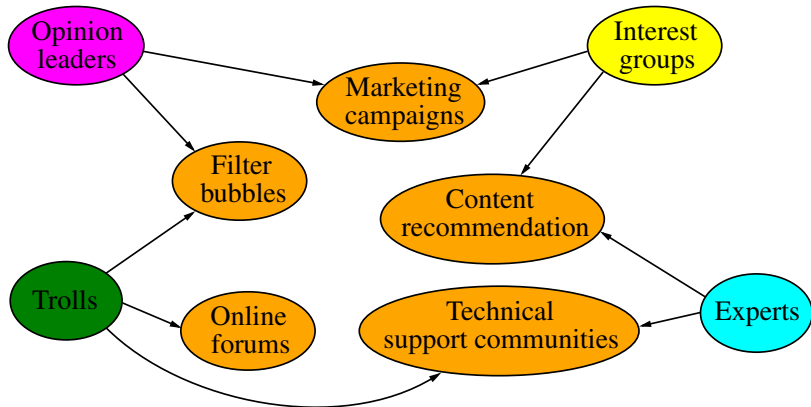
1 Introduction – Mechanism design challenges

2 Group identification

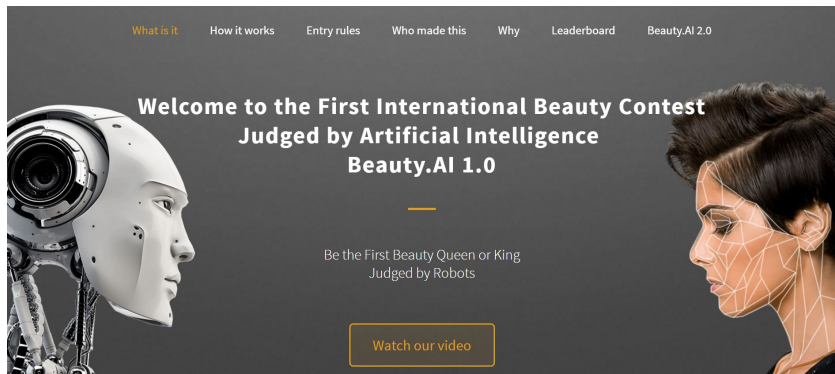
- Model framework
- Top Candidate algorithm
- Case study

3 Future research

Mechanism design challenges



Beauty.ai



- First beauty contest where entrants were judged entirely by an AI.
- 6,000 people from more than 100 countries submitted photos.
- Out of 44 winners only one had dark skin.

Group identification

The group identification problem was introduced by Kasher and Rubinstein to address a policy question related to Jewish identity.

Motivation

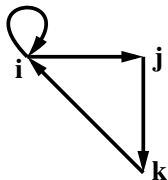
The "Law of Return" (1950) granted Jews the right of return and the right to live in Israel and to gain citizenship. In 1970, the right was extended to people of Jewish ancestry, and their spouses. A public debate has arisen concerning who is considered to be Jewish.

Description of the model

- Let $N = \{1, 2, \dots, n\}$ denote the set of individuals in the community.
- Based on the opinion of the individuals we would like to identify a certain subset of N .
- An opinion profile $P = (p_{ij})_{n \times n}$ is a matrix which contains the opinions, where $p_{ij} = 1$ if i believes that j belongs to the group, and $p_{ij} = 0$ otherwise.

Graph representation

$$N = \{i, j, k\} \quad P = \begin{matrix} & \begin{matrix} i & j & k \end{matrix} \\ \begin{matrix} i \\ j \\ k \end{matrix} & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{matrix}$$



Extension - non-elective members

We extend the framework of Kasher and Rubinstein in wan way: **we allow for some individuals to form opinion without being elective.**

- An examining committee is assembled and some persons are deemed unsuitable due to conflict of interest.
- A prize is distributed annually, and a person can not receive it twice.

Group identification problem

Definition

- A **group identification problem** Γ is a triple (N, P, X) consisting of the set of individuals N , the corresponding opinion profile P and a list X containing the non-elective members. The complement of X - the set of members who can be elected - is denoted by E .
- The set of group identification problems on N is denoted by \mathcal{G}^N .
- A **selection rule** is function $f : \mathcal{G}^N \rightarrow 2^E$ that assigns to each group identification problem a subset of the feasible set (i.e. the members of the group).

Axioms

Consensus

If everyone thinks that i belongs to the group (including i himself) then the selection rule should classify i as a group member. Conversely, if nobody thinks that i is a group member then the selection rule should not consider i as a group member.

Symmetry

Two individuals who are identical in the sense that

- everyone else thinks the same about them,
- they think the same about everyone else,
- and they treat each other in the same way,

are either both classified as group members or both are discarded.

Axioms

Monotonicity

If i is selected as a group member, then she will still be selected if more people believe that she is a group member (cf. positive responsiveness). If i is not selected as a group member, then she will still be omitted if less people think that she belongs to the group.

Independence[†]

Whether individual i is selected as a group member or not, should only depend on the views held about i .

†The original Kasher-Rubinstein axiom is weaker.

Self-identification rule

The Liberal Principle

If at least one individual thinks that he is a group member, then the selection rule should not return the empty set. If no one considers himself a group member the selection rule should not return the whole set.

Definition

The **self-identification rule** is the selection rule where $i \in f(P) \iff p_{ii} = 1$.

Theorem

A social rule satisfies consensus, symmetry, monotonicity, independence and the Liberal Principle if and only if it is the self-identification rule.

The problem with self-identification

- When the group characteristics depend on the inner beliefs of the individuals (e.g. ethnicity, religion), then self-identification works just fine.
- Obviously self-identification does not work when there is a more 'objective' criterion that determines who belongs to the group (e.g. celebrities, experts, trolls, etc.).
- Here we will focus our attention on expert groups.
- Potential applications of expert selection methods include identifying expert witnesses for jury trials, locating expertise in large companies, and creating shortlists of academic researchers or institutions.

Expert groups

Some expert groups can be found by competitions (e.g. the best chess players) others need a more delicate analysis. For instance we can not decide who is the best economist by competitions, but self-identification or simple majority voting will not suffice either. The key idea is the following.

- Experts and non-experts have different capabilities in identifying each other.
- Experts tend to identify each other better, while laypersons may rule out real experts and recommend dilettantes.

History

This idea has some history.

- A sociometric study by Seeley, J.R. (1949);
- Katz-index, Katz, L. (1953);
- Eigenvector centrality, Bonacich, P. and Lloyd, P. (2001);
- PageRank, Page, L and Brin, S. and Motwani, R. and Winograd T (1999).

Further notation

- If $p_{ij} = 1$ then we say that i **recommends** j or that j is a **candidate** of i .
- We denote by $N(i)$ the neighbours of i , i.e. the set of individuals who according to i 's opinion belong to the group.
- The supporters of i , the individuals who believe that i is a group member, is denoted by $B(i)$.

We allow for i to form an opinion about herself, that is, $N(i)$ and $B(i)$ may contain i .

Stable set

The stable set is the largest such set $S \subseteq N$ for which the following two requirements hold:

- 1) $i \in S \Rightarrow N(i) \subseteq S$, that is, the candidates of an individual in the set also belong to the set;
- 2) $i \in S \Rightarrow \exists j \in S$ such that $i \in N(j)$, that is, each individual in the set is supported by somebody in the set.

Remarks

- This stability notion is similar in concept to the von Neumann-Morgerstern stable set, both requiring inner and external stability.
- In standard social choice theory the alternatives are compared with a binary relation that is based on some kind of dominance notion.
- In the group identification framework recommendations carry a positive meaning (i.e. *not* a dominance relation).

Toward the core of the stable set

- The stable set is a large group which does not necessarily consists solely of experts.
- The individuals that are not in the stable set – by our argument – cannot be possibly experts.
- To locate the **core** of the stable set let us restrain the number of recommendations a person can make.

Proposed algorithm to find the core

- We ask each individual to **nominate** one person, who is the most prominent candidate for the group.
- Then we successively remove from the set each individual who is not nominated by anyone.
- If an individual loses support because we removed each person who nominated her, then we remove her too.
- We repeat this until either the set becomes empty or each individual in the set is nominated by somebody in the set.

Qualifiers

Definition

A function $Q : \mathcal{G}^N \rightarrow 2^N$ is called a **qualifier** if it satisfies the following two conditions

- $Q(i) \subseteq N(i)$ for all $i \in N$ and
- $Q(S) = \cup_{i \in S} Q(i)$ for any $S \subseteq N$.

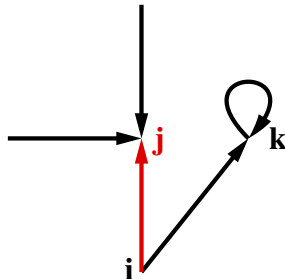
We say that i **nominates** j **under** Q if $j \in Q(i)$.

- Qualifiers serve as filters, they narrow down the possible group members.
- The set $Q(S)$ collects those individuals who are nominated by at least one person in S .
- Qualifiers – unlike to selection rules – may nominate non-elective members as well.

Nomination

Definition

We say that j is a **top candidate** of i if j has the most recommendations among the candidates of i . In case of a tie, when a person has more than one top candidate, we allow her to nominate all of them. The set of top candidates of individual i is denoted by $TC(i)$.



Top Candidate algorithm

(Initialization) $I_0 = N, k = 0$

while ($I_k \neq I_{k-1}$ or $I_k \neq \emptyset$)

{

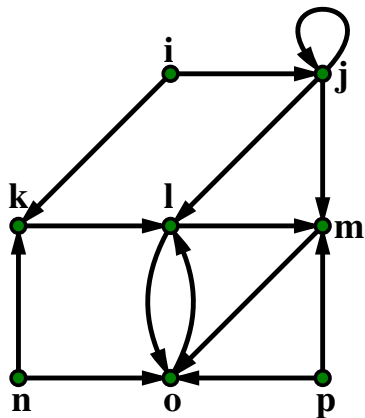
$k := k + 1$

 Let $I_k = \{j \in I_{k-1} \mid j \text{ is a top candidate for some } j' \in I_{k-1}\}$

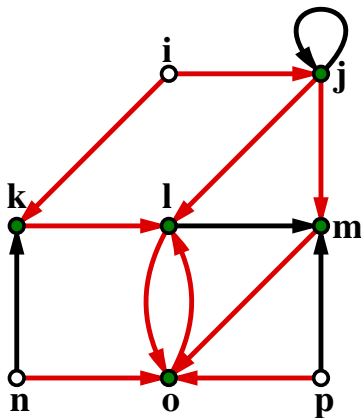
}

(Output) I_k

Start

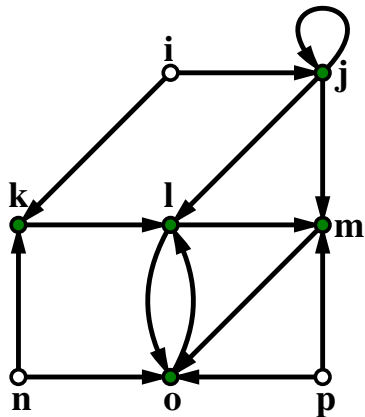


End of the 1st iteration

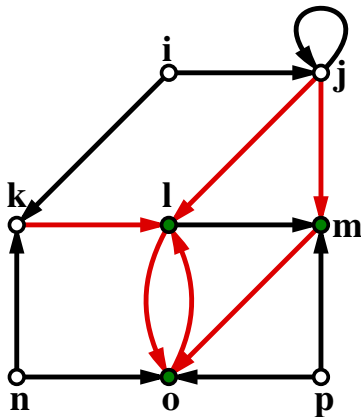


Green nodes: selected members
 Red arcs: Top candidate relation

Start of the 2nd iteration



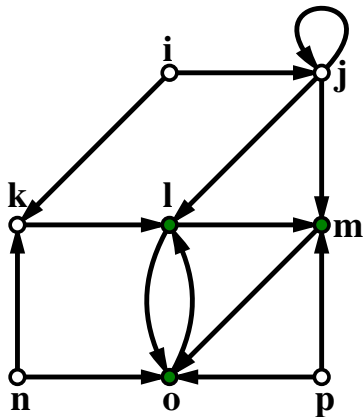
End of the 2nd iteration



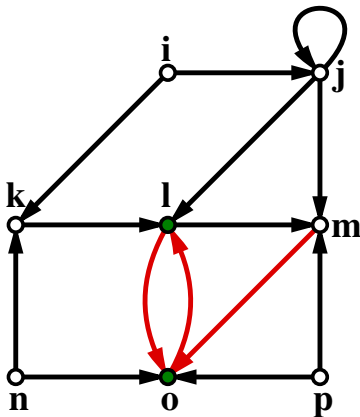
Green nodes: selected members

Red arcs: Top candidate relation

Start of the 3rd iteration



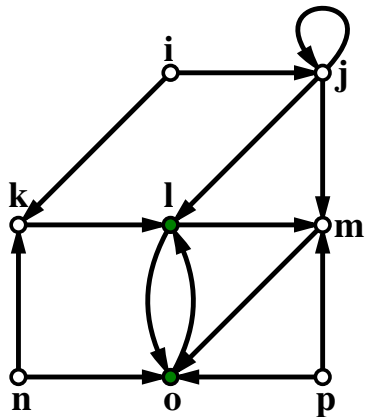
End of the 3rd iteration



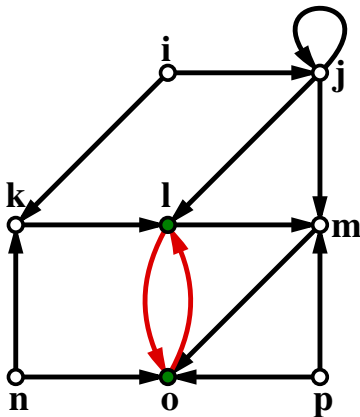
Green nodes: selected members

Red arcs: Top candidate relation

Start of the 4th iteration



Stop



Green nodes: selected members

Red arcs: Top candidate relation

Proposed axioms I.

Henceforward we will use the following notation $\Gamma_\emptyset = (N, P, \emptyset)$.

Weak Axiom of Revealed Preference

We say that a rule f satisfies **weak axiom of revealed preference** (WARP) if $f(\Gamma) = f(\Gamma_\emptyset) \setminus X$ for any $\Gamma = (N, P, X)$

- WARP implies that the selection rule does not distinguish between the opinion of the elective and excluded members.
- WARP is a standard axiom which is used e.g. in the extension of Arrow's Impossibility Theorem to choice sets.
- We will only need WARP to simplify the stability axiom, the characterization holds without it.

Proposed axioms II.a

(Strong) stability

Let $\Gamma = (N, P, X)$ be a GIP, Q a qualifier and f a selection rule. Furthermore let

$$X' \stackrel{\text{def}}{=} f(\Gamma_\emptyset) \setminus f(\Gamma).$$

Then we say that f is **stable with respect to** Q if

$$Q(f(\Gamma) \cup X') \subseteq f(\Gamma) \cup X' \text{ for all } \Gamma \in \mathcal{G}^N.$$

We say that f is **strongly stable with respect to** Q if

$$Q(f(\Gamma) \cup X') = f(\Gamma) \cup X' \text{ for all } \Gamma \in \mathcal{G}^N.$$

Proposed axioms II.b

(Strong) stability

Let $\Gamma = (N, P, X)$ be a GIP, Q a qualifier and f a selection rule that satisfies WARP. Then we say that f is **w-stable with respect to Q** if

$$Q(f(\Gamma_\emptyset)) \subseteq f(\Gamma_\emptyset) \text{ for all } \Gamma \in \mathcal{G}^N. \quad (1)$$

We say that f is **w-strongly stable with respect to Q** if

$$Q(f(\Gamma_\emptyset)) = f(\Gamma_\emptyset) \text{ for all } \Gamma \in \mathcal{G}^N. \quad (2)$$

- (1) The nominees of any expert should be included in the group of experts.
- (2) Each expert should be nominated by someone from the group of experts.

Proposed axioms III.

Exhaustiveness

For any group identification problem $\Gamma = (N, P, X)$ we define $\Gamma' = (N, P, X')$ to be the problem derived from Γ by setting $X' = X \cup f(\Gamma)$. We say that a rule f is **exhaustive** if $f(\Gamma') = \emptyset$ for each $\Gamma \in \mathcal{G}^N$.

- The exhaustiveness axiom requires from the selection rule to find every relevant participant.
- If, after excluding $f(\Gamma)$, the selection rule finds new experts, then the rule is not exhaustive – these individuals should have been included in the original group.
- WARP implies exhaustiveness.

Proposed axioms IV.

Since the rule that assigns the empty set for each GIP is both strongly stable and exhaustive, we need some kind of existence axiom as well.

Decisiveness

We call a subset of the individuals $C \subseteq N$ a **TC-component** if the top candidates of the individuals in C equals to C . A rule f satisfies **decisiveness** if $f(\Gamma) \neq \emptyset$ whenever there is a TC-component which has at least one elective member.

- Decisiveness is a weak requirement which ensures the non-emptiness of the solution set under general conditions.
- Note that the axiom does not require from the rule to select a TC-core member.
- For instance each GIP, where every individual has at least one recommendation, has a TC-component.

Characterization

Theorem

A selection rule satisfies strong stability with respect to the Top Candidate relation, exhaustiveness and decisiveness if and only if it is the TC-algorithm.

Groups vs. Centrality scores

- Centrality measures output a vector of real numbers which signifies the importance of the individuals, while our algorithm produces a list of individuals who are deemed important.
- A similar list can be obtained with centrality measures by setting a limit and declaring every individual important whenever his or her score is above the limit.
- However choosing the limit *ex ante* could lead to an arbitrary result, while setting the limit *a posteriori* is inherently biased by subjective elements.

Relaxing the Top Candidate relation

- The Top Candidate relation might be too strict in some cases.
- In the Relaxed TC-algorithm individuals nominate their 'best' $X\%$ of candidates (but at least one).
- Suppose Alice's top candidate is Bob who has 100 recommendations. Another candidate of Alice, Eve has 97. In the 3-Relaxed TC-algorithm Alice will also nominate Eve, since the difference between the number of individuals who recommend Bob and Eve is not more than 3%.

Relaxed Top Candidate algorithm

Definition

We say that j is a TC^X candidate of i , if $j \in N(i)$ and

$$|B(TC(i))| \left(1 - \frac{X}{100}\right) \leq |B(j)|.$$

The set of TC^X candidates for individual i is denoted by $Q_X(i)$

- Observe that $TC(i) \subseteq TC^X(i)$ for any $X \in [0, 100]$, and $TC^X(i) \subseteq TC^Y(i)$ if $X \leq Y$.
- In particular Q_0 will always yield the TC-core, while Q_{100} will result in the stable set.

Case study

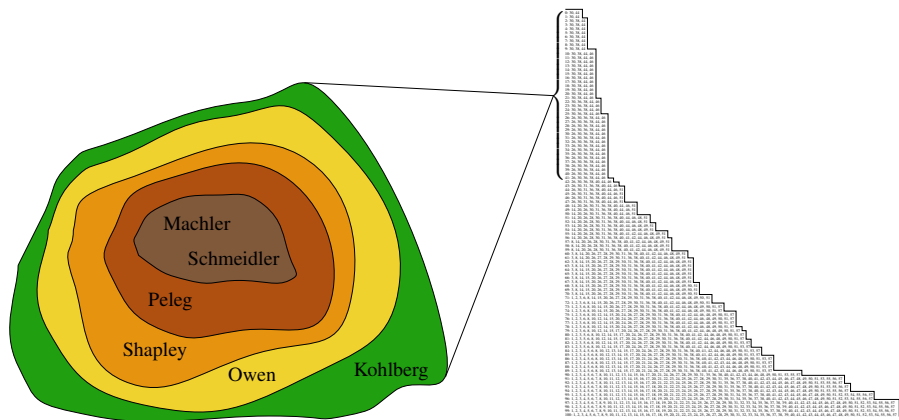
- A citation analysis of 88 articles of 57[†] authors focusing on the theoretical development of nucleolus.
- The citation graph has 57 nodes (authors) and 937 arcs (references).
- The opinion matrix was formed on the basis of the bibliography section of the articles. If author x cited author y in any of the reviewed papers then p_{xy} was set to 1.
- Disclaimer: this case study was designed to demonstrate the applicability of the TC-algorithm and not to realistically assess the contributions to this research topic.

† Some of the coauthors of these papers were omitted in the analysis.

Topographic map of the stable set

ID	Author	# of art.	# of ref.	Betweenness	Closeness	PageRank
1	A, H.	2	16	9.736	0.0119	1.036
2	A, J.	4	16	28.724	0.0120	1.106
3	A, R, J.	1	20	17.450	0.0114	0.908
4	B, R.	2	9	14.551	0.0119	1.041
5	D, X.	2	11	12.975	0.0120	1.073
6	D, J.	3	16	44.789	0.0127	1.235
10	F, U.	4	15	18.626	0.0122	1.107
12	F, V.	3	14	21.818	0.0116	1.006
14	G, D.	4	26	26.049	0.0127	1.203
15	G, F.	1	19	11.468	0.0119	1.041
17	H, H.	2	12	7.845	0.0112	0.886
20	H, G.	2	26	14.857	0.0122	1.103
24	K, W.	6	14	15.422	0.0120	1.077
26	K, E.	2	37	64.942	0.0133	1.372
27	K, A.	1	20	19.432	0.0110	0.821
28	K, J.	4	27	18.258	0.0127	1.194
29	L, S, C.	1	20	15.377	0.0111	0.877
30	Maschler, M.	6	50	159.629	0.0169	1.888
31	M, N.	3	29	35.683	0.0125	1.187
36	O, G.	2	39	68.788	0.0137	1.422
38	Peleg, B.	2	49	135.792	0.0159	1.759
40	P, J.	8	32	105.505	0.0152	1.655
41	R, T, E, S.	3	25	38.708	0.0132	1.305
42	R, H.	4	25	41.420	0.0132	1.308
44	Schmeidler, D.	1	55	208.239	0.0175	1.942
46	Shapley, L, S.	1	50	148.548	0.0164	1.821
48	S, A, I.	1	27	0.228	0.0097	0.423
49	S, T.	6	26	36.310	0.0119	1.087
50	S, P.	6	18	29.035	0.0118	1.050
51	T, S.	6	31	97.916	0.0149	1.614

Topographic map of the stable set



TC-core: Schmeidler, Maschler

X-Relaxed-core: Peleg ($X=2$), Shapley ($X=10$), Owen ($X=22$), Kohlberg ($X=26$)

Justification

Note that any citation analysis is inevitably biased by the "arrow of time" – older papers tend to have more citations.

- The outcome of the TC-algorithm is backed up by the three classic centrality measure.
- Schmeidler was who introduced the nucleolus. Maschler, Peleg and Shapley are the authors of one of the most influential papers in the subject, "Geometric Properties of the Kernel, Nucleolus, and Related Solution Concepts".
- Owen provided an LP to compute the nucleolus. Kohlberg introduced a criterion, which can be used to verify whether an allocation is the nucleolus or not.

Future research

- In progress: a ranking of game theory research institutions based on the citations they received between 2001-2011.
- Testing the TC-algorithm on sociometric data.
- Next project: identifying trolls and abusers.

News portal's dilemma

- A Hungarian news portal reported on a real estate website which had a questionable business practice.
- Some readers criticised the real-estate website in the comment section of the news article in derogatory terms.
- The company operating the website sued the news portal, which immediately deleted the comments.
- This was not enough, Hungarian courts found the news portal liable.
- The news portal appealed to the European Court of Human Rights and won (decision 22947/13).
- Still the news portal disallowed commenting on all of its articles.

News portal's dilemma

- The court decision cleared the news portal of blame, but did not resolve the issue. News portals might still be held accountable for their users' comments.
- A possible solution is using moderators to vet comments (high costs).
- Other is disallowing (anonymous) comments (decreases reader involvement).
- Court decision should endorse an algorithmic solution together with a notice-and-take-down system (a good algorithm is needed).

Issues to be solved

- Trolls may coordinate their attack
- Intense debate is not trolling (Trump vs. Clinton followers)
- Errors of first (not banning a troll) and second kind (banning an innocent)

Literature

- Bonacich, P. Lloyd, P. (2001): *Eigenvector-like measures of centrality for asymmetric relations*. Social Networks 23(3):191-201
- European Court of Human Rights (2016): Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary - 22947/13 Judgment 2.2.2016 [Section IV]
- Kasher, A. Rubinstein, A. (1997): *On the question "Who is a J?", a social choice approach*. Logique et Analyse, Vol. 160 pp. 385-395.
- Katz, L. (1953): *A new status index derived from sociometric index*. Psychometrika 18:39-43
- Levin, S. (2016): *A beauty contest was judged by ai and the robots didn't like dark skin*. The Guardian, San Fransisco (US)
- Page, L. Brin, S. Motwani, R. Winograd, T. (1999): *The PageRank citation ranking: Bringing order to the web*
- Seeley, J.R. (1949): *The net of reciprocal influence; a problem in treating sociometric data*. Canadian Journal of Psychology 3:234-240