How to identify experts in the community?

Balázs Sziklai

XXXII. Magyar Operációkutatás Konferencia, Cegléd

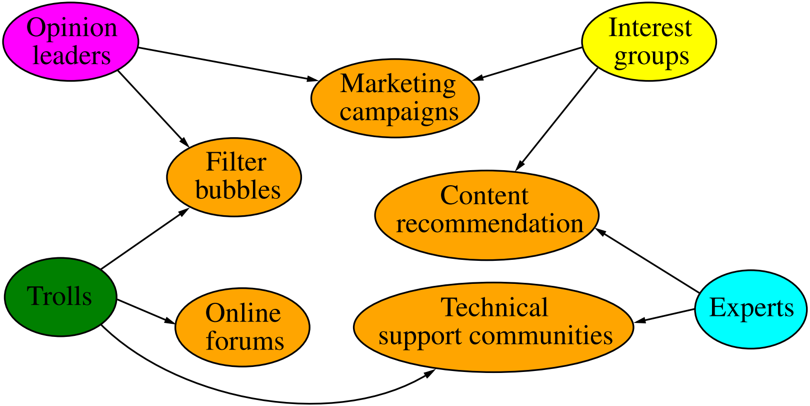
e-mail: sziklai.balazs@krtk.mta.hu

2017. 06. 15.

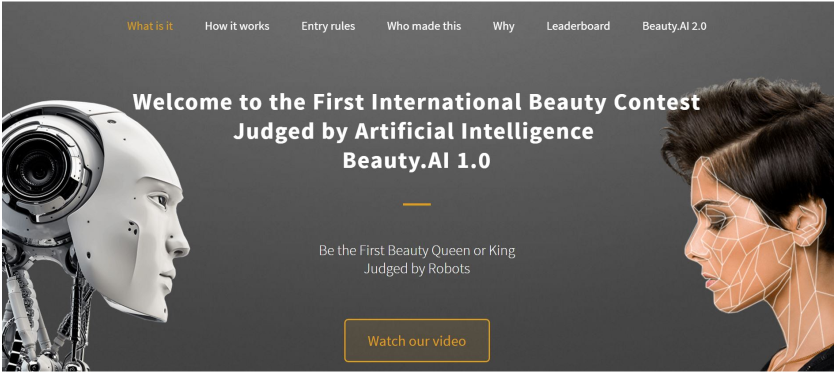
1. **Introduction – Mechanism design challenges**

2. **Group identification**
   - Model framework
   - Top Candidate algorithm
   - Case study

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

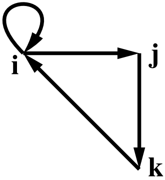
The "Law of Return" (1950) granted Jews the right of return and the right to live in Israel and to gain citizenship. In 1970, the right was extended to people of Jewish ancestry, and their spouses. A public debate has arisen concerning who is considered to be Jewish.

Description of the model

- Let $N = \{1, 2, \ldots, n\}$ 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

$$N = \{i, j, k\} \qquad P = \begin{array}{c} \\ i \\ j \\ k \end{array} \begin{array}{c} \begin{array}{ccc} i & j & k \end{array} \\ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$



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 $f : \mathcal{G}^N \to 2^E$ that assigns to each group identification problem a subset of the feasible set (i.e. the members of the group).

Self-identification rule

Definition

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

Theorem (Kasher&Rubinstein [1997])

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.

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 $Q : \mathcal{G}^N \to 2^N$ 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 $S \subseteq N$.

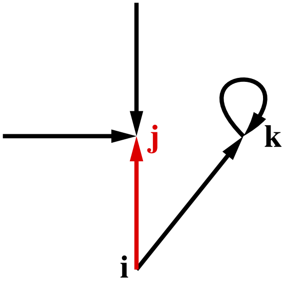
We say that i **nominates** j **under** Q if $j \in Q(i)$.

- 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 $TC(i)$.



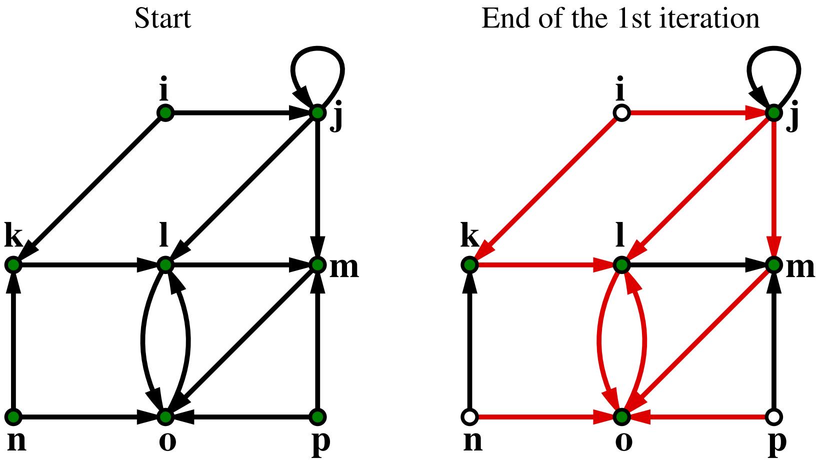
Top Candidate algorithm

Core Selection Algorithm (w.r.t. Q)

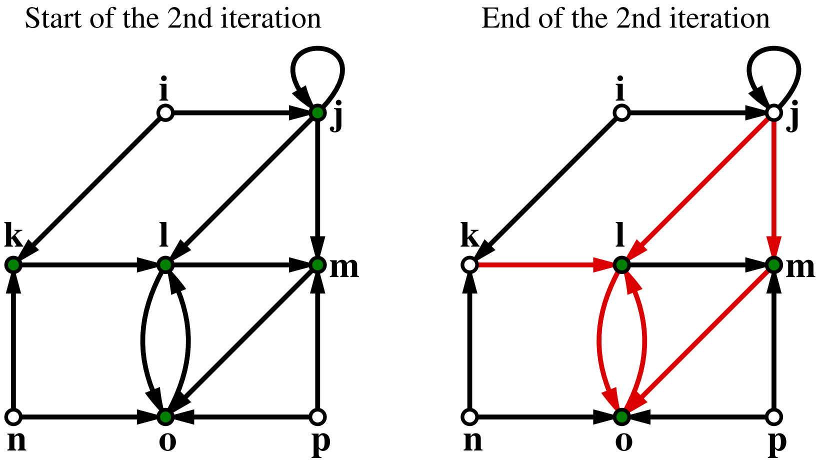
(Input) N, P, X

(Initialization) $I_0 = N,\ k = 0$
while $(I_k \neq I_{k-1}$ or $I_k \neq \emptyset)$
$\{$
$k := k + 1$
$I_k := \{j \in I_{k-1} \mid j \in Q(j')$ for some $j' \in I_{k-1}\}$
$\}$
(Output) $I_k \setminus X$

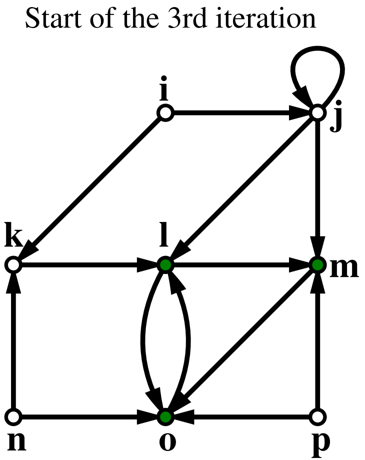
If Q is set to Q_T we refer to the above method as the *top candidate*- or shortly *TC-algorithm* and the obtained set as the *TC-core*.



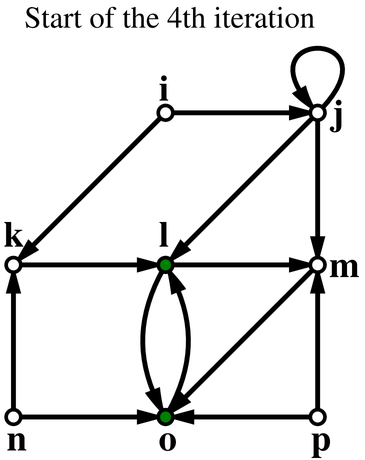
Green nodes: selected members
Red arcs: Top candidate relation

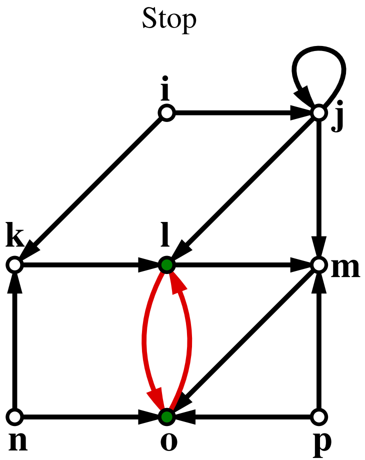


Green nodes: selected members
Red arcs: Top candidate relation



Green nodes: selected members
Red arcs: Top candidate relation





Green nodes: selected members
Red arcs: Top candidate relation

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' $\alpha\%$ 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^α candidate of i, if $j \in N(i)$ and

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

The set of TC^α candidates for individual i is denoted by $Q_\alpha(i)$

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

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

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

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

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

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

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

(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.

Definition

We call a subset of the individuals $C \subseteq N$ a *stable component with respect to* Q if $Q(C) = C$.

α-Decisiveness

A rule f satisfies *α-decisiveness* if $f(\Gamma) \neq \emptyset$ whenever there is a stable component with respect to Q_α which has at least one elective member. A 0-decisive rule is simply called *decisive*, while a 100-decisive rule is called *permissive*.

- The parameter α expresses how permissive the selection rule is.
- Note that the axiom does not require from the rule to select a core member.
- For instance each GIP, where every individual has at least one recommendation, has a TC-component.

Characterization

Let f_α be the selection rule that returns the elective members of the TC^α-core, that is

$$f_\alpha(\Gamma) = C_\alpha.$$

Theorem

A selection rule satisfies strong stability with respect to Q_α, exhaustiveness and α-decisiveness if and only if it is $f_\alpha(\Gamma)$.

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.

Case study

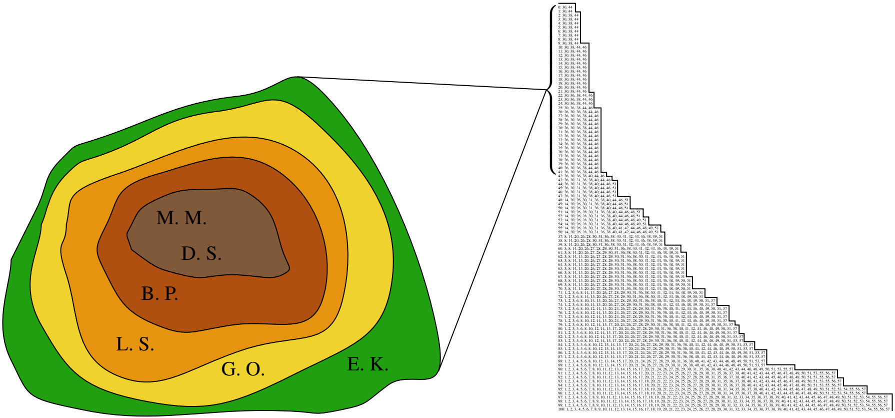
- A citation analysis of 88 articles of 57[†] authors focusing on a cooperative game theoretical topic.
- The citation graph has 57 nodes (authors) and 937 arcs (references).
- The opinion matrix was formed on the basis of the bibliography section of the articles. If author x cited author y in any of the reviewed papers then p_{xy} was set to 1.

† 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	M, 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	P, 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	S, D.	1	55	208.239	0.0175	1.942
46	S, 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: D. S., M. M.

X-Relaxed-core: B. P. (X=2), L. S. (X=10), G. O. (X=22), E. K. (X=26)

Literature

- Bonacich, P. Lloyd, P. (2001): *Eigenvector-like measures of centrality for asymmetric relations.* Social Networks 23(3):191-201
- 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