MAGYAR TUDOMÁNYOS AKADÉMIA
Közgazdaság- és Regionális Tudományi Kutatóközpont

MTA KRTK KTI

Centre for Economic and Regional Studies
HUNGARIAN ACADEMY OF SCIENCES

MŰHELYTANULMÁNYOK DISCUSSION PAPERS

MT-DP – 2015/49

On how to identify experts in a community

BALÁZS SZIKLAI

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
HUNGARIAN ACADEMY OF SCIENCES BUDAPEST, 2015

Discussion papers
MT-DP – 2015/49

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

On how to identify experts in a community

Author:

Balázs Sziklai
junior research fellow
Momentum Game Theory Research Group
Institute of Economics - Centre for Economic and Regional Studies
Hungarian Academy of Sciences
e-mail: sziklai.balazs@krtk.mta.hu

September 2015

ISBN 978-615-5594-14-4
ISSN 1785 377X

On how to identify experts in a community

Balázs Sziklai

Abstract

The group identification literature mostly revolves around the problem of identifying individuals in the community who belong to groups with ethnic or religious identity. Here we use the same model framework to identify individuals who play key role in some sense. In particular we will focus on expert selection in social networks. Ethnic groups and experts groups need completely different approaches and different type of selection rules are successful for one and for the other. We drop monotonicity and independence, two common requirements, in order to achieve stability, a property which is indispensable in case of expert selection. The idea is that experts are more effective in identifying each other, thus the selected individuals should support each others membership. We propose an algorithm based on the so called top candidate relation. We establish an axiomatization to show that it is theoretically well-founded. Furthermore we present a case study using citation data to demonstrate its effectiveness. We compare its performance with classical centrality measures.

Keywords: Group identification, Expert selection, Stability, Citation analysis, Nucleolus

JEL classification: D71

Acknowledgement:

Research was funded by OTKA grants K109354 and K108383 and by the Hungarian Academy of Sciences under its Momentum Programme (LD-004/2010).

Hogyan azonosítsunk szakértőket egy közösségben?

Sziklai Balázs

Összefoglaló

A csoportidentifikációval foglalkozó irodalom többnyire etnikai vagy vallási csoportok azonosítását érintő kérdéseket vizsgál. Műhelytanulmányunkban arra használjuk a modellt, hogy azonosítsuk a közösségben valamilyen szempontból kulcsszerepet betöltő szereplőket. Ezen belül is a szakértők társadalmi hálózatokon belüli azonosítására összpontosítunk. Etnikai csoportok és szakértő csoportok teljesen különböző megközelítést igényelnek, más kiválasztási szabály lesz sikeres az egyik és a másik esetében. Elhagyjuk a kiválasztási szabályok két gyakran alkalmazott tulajdonságát – a monotonitást és a függetlenséget – annak érdekében, hogy biztosítsuk a csoport stabilitását. Ez utóbbi elengedhetetlennek bizonyul a szakértők meghatározásakor. A fő gondolat az, hogy a szakértők sikeresebben azonosítják egymást, tehát a megalakuló csoportnak belső támogatottsággal kell rendelkeznie. Bemutatunk egy algoritmust, ami az ún. csúcsjelölt-kiválasztáson alapul. Az így nyert kiválasztási szabályt egy axiomatizáció segítségével elméletileg is megalapozzuk. A módszer hatékonyságát egy citációs adatbázison alapuló esettanulmányon is szemléltetjük. Az algoritmus teljesítményének kiértékeléséhez a kapott eredményeket összevetjük azokkal, amelyeket a klasszikus centralitás mértékek jósolnak.

Tárgyszavak: csoportidentifikáció, szakértő-kiválasztás, stabilitás, citációs elemzés, nukleolusz

JEL kód: D71

On how to identify experts in a community

Balázs Sziklai*

September 12, 2015

Abstract

The group identification literature mostly revolves around the problem of identifying individuals in the community who belong to groups with ethnic or religious identity. Here we use the same model framework to identify individuals who play key role in some sense. In particular we will focus on expert selection in social networks. Ethnic groups and experts groups need completely different approaches and different type of selection rules are successful for one and for the other. We drop monotonicity and independence, two common requirements, in order to achieve stability, a property which is indispensable in case of expert selection. The idea is that experts are more effective in identifying each other, thus the selected individuals should support each others membership. We propose an algorithm based on the so called top candidate relation. We establish an axiomatization to show that it is theoretically well-founded. Furthermore we present a case study using citation data to demonstrate its effectiveness. We compare its performance with classical centrality measures.

Keywords and phrases: Group identification, Expert selection, Stability, Citation analysis, Nucleolus

JEL-codes: D71

1 Introduction

The group identification literature has been focusing primarily on social categories such as ethnicity and religion. The original model of Kasher and

*Research was funded by OTKA grants K109354 and K108383 and by the Hungarian Academy of Sciences under its Momentum Programme (LD-004/2010).

Rubinstein (1997) stemmed from questions related to Jewish identity, while Miller (2008) reports on federal policy regulating racial data collection in the US. They argue that self-identification is the only conceptually sound selection rule. Self-identification indeed seems to be the good choice when the group's characteristics depend on the inner beliefs of the individuals, however, it does not fare so well when there are more objective traits which define who belongs to the group. For instance determining who is the best chess player is an altogether different problem. The latter kind of questions are usually decided by competitions. There is an extensive literature on tournament solutions, for a comprehensive review see (Laslier, 1997).

Samet and Schmeidler (2003) consider a broader spectrum of group identification problems ranging from issues that can be decided by the individual, such as who can read a book, to problems which need some kind of social consent, e.g. electing a candidate. They view it as the confrontation of two democratic principles: liberalism (or individualism) and majoritarianism. They propose a family of rules characterized by three axioms, symmetry, monotonicity and independence[1]. The relationship of the latter two are also studied by Ju (2010) and Çengelci and Sanver (2010).

Some groups, however, cannot be captured by voting rules which satisfy independence or monotonicity. Consider the problem of identifying underground[2] music bands. A band which is referred too many times is by definition mainstream, and the individuals who refer to it have false perception regarding its popularity. Politicians who will achieve a surprisingly good result on the next election compose another group which is impossible to capture with monotonic voting rules. If everyone believes that a politician will produce a good result in the upcoming election, then his success is hardly surprising.

Independence is also violated in some cases. The problem of identifying experts in a community incorporates both subjective and objective elements. We can not decide who is the best economist by competitions, but self-appointment or simple majority voting will not suffice either. The latter one fails because 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.

We will translate the recommendations of the individuals into a directed graph network. Experts in the graph are nodes with certain desirable features. The general problem, measuring the significance of some nodes in a

[1]The formalization of this three axiom substantially differs from how they were established in (Kasher and Rubinstein, 1997). In particular monotonicity and independence were much weaker. Note that Miller (2008) refers to symmetry as anonymity.

[2]Here 'underground' refers to the popularity of the band and not to the music style.

network, occurs in various field, such as Computer science, Chemistry or Biology. Various methods were invented in parallel to quantify the importance of these nodes. Boldi and Vigna (2014) offers a axiomatic overview on the most commonly applied centrality measures.

In this paper we focus on the axiomatic foundation of expert selection. We argue that stability is a key component of any solution. That is, the selected group should consider each of its members as experts and no one else outside the group. Depending on the qualification criterion we distinguish between different types of stability. We provide an iterative algorithm that is strongly stable with respect to the so called top candidate relation. Top candidates of an individual are persons who are approved by both the community and by the individual in question. Unlike in (Dimitrov et al., 2007) where the procedure starts from a small set and gradually extends the group until a certain condition holds, we start with the whole community and shrink the group size until stability is met.

We conclude with a case study based on citation data. A citation can be considered as a recommendation made by one author to another, which fits well with the group identification framework. We chose the nucleolus – a rather specific, but still extensive research topic in cooperative game theory – as the subject of our analysis.

Ranking scientific researchers based on their productivity and the recognition of their work is a popular topic with many applications, such as performance evaluation, recruitment, or handing out of a prize or a grant. The literature mostly focuses on developing and analysing indexes such as the h-index (Hirsch, 2005) and the g-index (Egghe, 2006). Our method provides a complementary tool for this task. The main difference between the classical metrics and our method is that instead of coming up with numbers that represent the importance of the individuals our algorithm selects a few individual who are deemed important.

Potential applications of the proposed model include content recommendation (Amatriain et al., 2009; Carchiolo et al., 2015), identifying professionals in community technical supports (Pal and Konstan, 2010) and locating competencies and expertise in large enterprises (John and Seligmann, 2006).

2 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.

If $p_{ij} = 1$ then we say that i *recommends* j. We assume that everyone states his or her true preference, the opinion matrix is not affected by modesty, envy or any strategic behaviour.

It is natural to think of P as the adjacency matrix of a directed graph, whose node and arc set are the individuals and their recommendations respectively. 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[3], that is, $N(i)$ and $B(i)$ may contain i.

We extend the model of Kasher and Rubinstein in one way: we allow for some individuals to form opinion without being elective. That is, some individuals cannot be chosen as a group member. This is quite natural in some applications. Suppose, for instance, that there is a prize which is awarded annually to the best economist. Individuals who won the prize on a previous occasion can not be chosen again. Their opinion however matters. The same problem happens when an examining committee is assembled and some persons are deemed unsuitable due to conflict of interest. For example an editor may not like to hand over a manuscript to a former coauthor of the submitter, but he may inquire his opinion about the referee selection. To ensure that every relevant information is encompassed in the decision, we allow the non-elective members to form opinion and also for others to form opinion about them.

A group identification problem (shortly GIP) Γ is triple (N, P, X) consisting the set of individuals N, the corresponding opinion profile P and a list X containing the non-elective members. The complement of X - the members who can be elected - are denoted by E. The GIP $(N, P, \emptyset)$ where every individual is elective is of special importance and will be denoted by $\Gamma_\emptyset$. 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 a set of individuals (i.e. the members of the group) for each GIP. The most widely studied selection rule is the liberal rule (aka self-identification), denoted by L, which picks the elective individuals who consider themselves group members, formally $L(\Gamma) = \{j \in E \mid j \in N(j)\}$.

[3]We would like to avoid even the appearance that selecting experts is a 'man's job'. Thus we refer to individual i as 'she' and individual j as 'he'.

3 Proposed Axioms

Stability of a solution is a central concept both in mechanism design and game theory. Stability is also crucial for expert selection algorithms: the selected members must support each other's membership. Ultimately, the best judge of an expert is another expert.

This idea has some history. Samet and Schmeidler (2003) examine the problem of identifying Hobbits in the community. They introduce the so called *affirmative self-determination* axiom, which requires that Hobbits and only Hobbits determine who Hobbits are. However, they use it to characterize the liberal rule by applying standard axioms like monotonicity and independence. Miller (2008) proposes two new class of rules: agreement and nomination rules. Both of them – with different intensity – require from a group member to have inner support. The combination of these two rule types yields one-vote rules where the relevant set, which decides the group membership of an individual, consist of one person. Once again the liberal rule is characterized by using the so called separability axioms.

The stability requirement we propose has two aspects. First each member of the group must have some inner support. Secondly, if an individual is recognized as a group member then the persons he recommends are also potential members. To treat these conditions formally we introduce the notion of qualifiers.

Definition 1. Let $Q : 2^N \to 2^N$ be a selection criterion that assigns a set of individuals to any subset of N. A selection criterion Q 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. Note that qualifiers – unlike to selection rules – may nominate non-elective members as well. Let us clarify that the identification of a qualified persons is made by the aggregator and not by the individuals themselves. In this model we do not take into consideration whether an individual has a preference order on the other individuals or not.

We propose the following qualifier. Let each individual point to the person(s) among those he approves who are the most acknowledged in the community. The persons selected by i in this way are called the top candidates of i.

Definition 2. We say that j is a *top candidate* of i if $j \in N(i)$ and $|B(j)| \geq |B(j')|$ for any other $j' \in N(i)$.

Note that a person can have more than one top candidate. The set of top candidates for individual i is denoted by $Q_T(i)$, and we will use the following notation $Q_T(S) = \cup_{i \in S} Q_T(i)$. Now we are ready to formulate our stability axiom.

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 $Q(f(\Gamma) \cup X') \subseteq f(\Gamma) \cup X'$ for all $\Gamma \in \mathcal{G}^N$. We say that f is *strongly stable with respect to* Q if $Q(f(\Gamma) \cup X') = f(\Gamma) \cup X'$ for all $\Gamma \in \mathcal{G}^N$.

The set X' collects those members which would have been selected by f, were X not excluded in Γ. According to rule f only the individuals contained in $f(\Gamma) \cup X'$ may have relevant opinion. Stability requires that each nomination made by $f(\Gamma) \cup X'$ under the qualifier Q refers to someone inside the relevant group. In addition to this, strong stability requires that each of the selected individuals should be nominated by another group member.

The interpretation and the formal treatment of the stability axiom can be simplified if we restrict our attention to selection rules which do not distinguish between the opinion of the elective and excluded members.

Equal treatment of members: We say that a rule f satisfies *equal treatment of members* (ETM) if $f(\Gamma) = f(\Gamma_\emptyset) \setminus X$ for any $\Gamma = (N, P, X)$.

Note that ETM indeed simplifies the conditions of (strong) stability. Instead of $Q(f(\Gamma) \cup X') \subseteq f(\Gamma) \cup X'$ it is enough to require $Q(f(\Gamma_\emptyset)) \subseteq f(\Gamma_\emptyset)$ for all $\Gamma \in \mathcal{G}^N$ (substitute '=' instead of '$\subseteq$' in case of strong stability).

We argue that strong stability with respect to Q_T is a good selection requirement. Top candidates are approved by both the individual and the community. Strong stability with respect Q_T requires that

1. each expert should be a top candidate of at least one other expert, and
2. the top candidates of any expert should be also included in the group of experts.

In decentralized networks where there are many 'hubs' any reasonable selection algorithm will pick experts locally. For instance if the opinion graph consist of several disjoint or 'almost disjoint' components, it is more natural if the selection rule includes experts from each component. Suppose we would like to choose the economist of the year and we inquire the opinion of every economist. It is plausible that both micro- and macroeconomists will support someone among themselves. In such a heterogenous society it is better if we allow the selection rule to compose a mixed group. Then, after we narrowed down the number of possible recipients of the prize, we can decide the winner by qualitative analysis. From time to time a macroeconomist will recommend a microeconomist and vice versa. Thus we cannot decompose the opinion graph to disjoint components. That is, we should require more, than simply demanding from the selection rule to pick an expert from each component.

Exhaustiveness: Let $\Gamma = (N, P, X)$ be a group identification problem and f a selection rule. Let $\Gamma' = (N, P, X')$ be another 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.

Note that ETM implies exhaustiveness. Let $\Gamma_1 = (N, P, X_1)$ and $\Gamma_2 = (N, P, X_1 \cup f(\Gamma_1))$ where f is a rule that satisfies ETM. Then $f(\Gamma_1) = f(\Gamma_\emptyset) \setminus X_1$. We need to show that $f(\Gamma_2) = \emptyset$, indeed

$$f(\Gamma_2) = f(\Gamma_\emptyset) \setminus (X_1 \cup f(\Gamma_1)) = f(\Gamma_\emptyset) \setminus f(\Gamma_\emptyset) = \emptyset.$$

On the other hand there are rules which are exhaustive but does not satisfy ETM. It is easy to check that

$$r(\Gamma) = \begin{cases} N & \text{if } X = \emptyset, \\ \emptyset & \text{if } X \neq \emptyset. \end{cases}$$

is one such rule.

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.

TC-component existence: We call a subset of the individuals $C \subseteq N$ a *TC-component* if $Q_T(C) = C$. A rule f satisfies TC-component existence if $f(\Gamma) \neq \emptyset$ whenever P contains a top candidate component which has at least one elective member.

TC-component existence is a weak requirement which ensures the non-emptiness of the solution set under general conditions. For instance each GIP, where every individual has at least one recommendation, has a TC-component. This follows from the fact that any individual which has a recommendation has also a top candidate. In such cases there exist a cycle of top candidates, from which it follows that a TC-component exists.

4 The Expert Selection Algorithm

The method we propose is based on the same concept as the PageRank algorithm (Page et al., 1999). There is one fundamental difference between the two approaches. While the PageRank – similarly to other measures of centrality – outputs a vector of real numbers that describes the importance of the individuals, our algorithm produces a set of individuals who are deemed important. Similar results can be obtained by setting a limit and declare every individual important whenever his or her score is above the limit. However choosing the limit is not an easy task. An ex ante decision could lead to an arbitrary result, while setting the limit a posteriori is inherently biased by subjective elements, and the outcome might be viewed as prejudicial.

The algorithm proceeds as follows. Each individual points to its top candidate among his or her neighbours. Then the procedure is repeated with the individuals who received at least one nomination. During the iteration we do not restrict ourselves to the opinions of the nominees, each recommendation still counts. Hence the set of top candidates for any person remains unchanged all through. The algorithm stops when it produces the same set of players twice. Formally

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

We refer to the result of the algorithm as the *core of* Γ *with respect to* Q_T or simply the *TC-core* and denote it with $TC(\Gamma)$. Note that in the last step we make sure that $TC(\Gamma)$ does not contain any excluded members. As a direct consequence the TC-algorithm satisfies ETM.

Example 3. Consider the opinion profile depicted in Figure 1. The individuals make the following nominations: $Q_T(i) = Q_T(m) = \{j\}$, $Q_T(j) = Q_T(l) = Q_T(n) = \{m\}$, $Q_T(k) = \emptyset$. Only individuals j and m receive nominations, hence $I_1 = \{j, m\}$. Since $Q_T(m) = \{j\}$ and $Q_T(j) = \{m\}$, it follows that $I_2 = \{j, m\}$ too. Thus, the algorithm stops and concludes that the core experts are j and m.

Theorem 4. *A selection rule satisfies strong stability with respect to* Q_T*,*

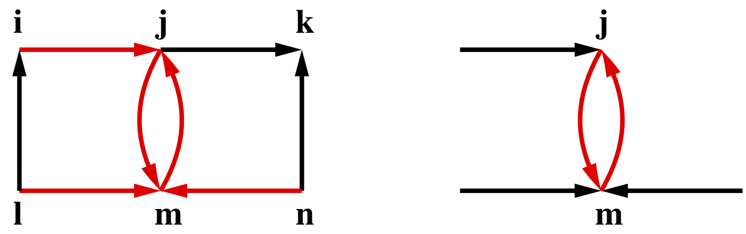

Figure 1: Graph representation of an opinion profile and the first (and last) iteration of the TC-algorithm. Red arcs indicate top candidate relation.

exhaustiveness and TC-component existence if and only if it is the TC-algorithm.

Proof. First we prove the *if* part. Let $\Gamma = (N, P, X)$ be an arbitrary GIP. If $I_k = \emptyset$ for some $k \geq 1$ then $TC(\Gamma) = \emptyset$ independently of the choice of X. In particular $TC(\Gamma_\emptyset) = \emptyset$, thus $Q_T(TC(\Gamma_\emptyset)) = Q_T(\emptyset) = \emptyset$. That is strong stability holds. If $I_{k-1} = I_k \neq \emptyset$ for some $k \geq 1$ then $TC(\Gamma_\emptyset) = I_k$ and

$$Q_T(TC(\Gamma_\emptyset)) = Q_T(I_k) = Q_T(I_{k-1}) = I_k = TC(\Gamma_\emptyset).$$

hence by ETM the TC-algorithm is strongly stable. Exhaustiveness again follows from ETM. Since I_k is derived independently of X, if X is set as I_k then the algorithm is forced to yield $\emptyset$. Finally note that each TC-component is included in each I_j set. Thus, if there is at least one TC-component with an elective member then the TC-algorithm will not result in the empty set.

Now we prove the *only if* part. Let f be a selection rule that satisfies strong stability, exhaustiveness and TC-component existence. No stable set (with respect to Q_T) contains individuals from $N \setminus I_1$ since these are not top candidates of anybody. Similarly no stable set contains individuals from $I_1 \setminus I_2$ as these are recommended only by individuals from $N \setminus I_1$. Consequently no stable set contains individuals from $I_{j-1} \setminus I_j$ for $1 \leq j \leq k$. Thus, if there exist $i \in f(\Gamma) \setminus I_j$ for some $0 \leq j \leq k$ then $f(\Gamma)$ is not strongly stable. This implies $f(\Gamma) \subseteq TC(\Gamma)$.

Let $\Gamma = (N, P, X)$ and let $i \in E$ be a member of a TC-component. Suppose that $i \notin f(\Gamma)$ and let $\Gamma' = (N, P, X \cup f(\Gamma))$. By TC-component existence $f(\Gamma') \neq \emptyset$. This contradict exhaustiveness of f. Hence f must contain every TC-component which has at least one elective member, that is $TC(\Gamma) \subseteq f(\Gamma)$. □

An interesting variant of the TC-algorithm can be derived by slightly altering its setup. One could argue that the opinion of those individuals who did not receive a nomination does not count. Thus at every iteration

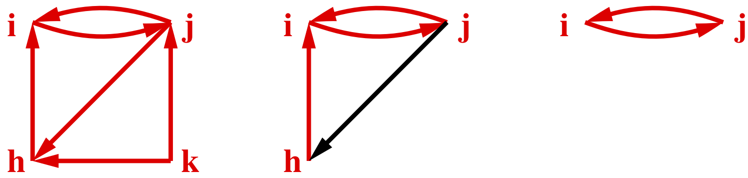

Figure 2: From left to right: Graph representation of the original opinion profile and the profiles of the restricted games which correspond to the first and second iteration of the eliminative TC-algorithm. Red arcs indicate top candidate relation.

we could just delete the nodes and arcs that belong to the non-nominated members. To capture this idea we need the concept of restricted game.

Definition 5. Let $\Gamma = (N, P, X)$ be a GIP, and $S \subset N$ a set of individuals. The restricted problem Γ_{S^-} is the GIP $(N \setminus S, P_{S^-}, X \setminus S)$ where P_{S^-} denotes the profile that is derived from P by deleting the rows and columns corresponding to the members of S.

The *eliminative top candidate algorithm* proceeds as follows.

(Initialization) $\Gamma_0 = (N, P, X), I_0 = N,\ k = 0\ S = \emptyset$
while ($I_k \neq I_{k-1}$ or $I_k \neq \emptyset$)
{
$k := k + 1$
$I_k := \{j \in I_{k-1} \mid j$ is a top candidate in Γ_{k-1} for some $j' \in I_{k-1}\}$
$S := N \setminus I_k$
$\Gamma_k := (N \setminus S, P_{S^-}, X \setminus S)$
}
(Output) $I_k \setminus X$

We refer to the result of the algorithm as the *eliminative core of* Γ *with respect to* Q_T or simply the *eliminative TC-core.*

Example 6. Consider the opinion profile depicted in Figure 2. The nominations are: $Q_T(h) = \{i\}$, $Q_T(i) = \{j\}$, $Q_T(j) = \{h, i\}$ and $Q_T(k) = \{h, j\}$. Since k is not nominated by anyone, his opinion is disregarded in the remaining part of the process. This affects the nominations as follows: $Q_T(h) = \{i\}$, $Q_T(i) = \{j\}$, $Q_T(j) = \{i\}$. Now h is not nominated anymore and he is dropped from the selection process. Finally i an j are chosen as eliminative core members.

Note that in this case the eliminative TC-algorithm resulted in a smaller set than the TC-core as the latter would have included h among the group as well. In general, however, the core and the eliminative core has no obvious relation. We will see in Section 6 a case where the eliminative core is a superset of the core.

Finally we note that both versions of the TC-algorithm are fast. If the community consist of n individuals the algorithms need to perform at most $\mathcal{O}(n^2)$ operations.

5 Axiomatic analysis

Kasher and Rubinstein (1997) define five basic axioms which characterize the liberal rule: consensus, symmetry, monotonicity, independence and the Liberal Principle. Sung and Dimitrov (2005) showed that these five axioms are not independent and symmetry, independence and the Liberal Principle are already enough for the characterization.

The Top Candidate algorithm satisfies consensus and symmetry, the two most basic requirements, that can be imposed on selection rules.

Definition 7. A selection rule f satisfies *consensus* if for any $i \in E$

- $B(i) = N \Rightarrow i \in f(\Gamma)$ and
- $B(i) = \emptyset \Rightarrow i \notin f(\Gamma)$

for any opinion profile P.

Definition 8. We say that two elective individuals i and j are symmetric in profile P if,

- everyone else thinks the same about them
- they think the same about everyone else
- i thinks that he belongs to the group iff j thinks that he belongs to the group
- i and j think the same about each other

A selection rule is *symmetric* if for every pair of such individuals it is true that $i \in f(\Gamma) \Leftrightarrow j \in f(\Gamma)$.

If i and j are symmetric in profile P then $i \in Q_T(k) \Leftrightarrow j \in Q_T(k)$ for any $k \in N \setminus \{i, j\}$ and $i \in Q_T(i) \Leftrightarrow j \in Q_T(j)$. By strong stability every expert is qualified by another expert, hence if $i, j \in E$ are symmetric and $i \in f(\Gamma)$ then this implies that $j \in f(\Gamma)$.

Definition 9. Let $\Gamma = (N, P, X)$ and $\Gamma' = (N, P', X)$ be two GIPs where P and P' are identical profiles except that $p_{ij} = 0$ and $p'_{ij} = 1$ for some $i, j \in N$. A selection rule is called *monotonic* if $i \in f(\Gamma) \Rightarrow i \in f(\Gamma')$ for any such pair of problems $\Gamma, \Gamma' \in \mathcal{G}^N$. A selection rule is called *group monotonic* if $f(\Gamma) \subseteq f(\Gamma')$ for any such pair of problems $\Gamma, \Gamma' \in \mathcal{G}^N$.

Group monotonicity was introduced by Samet and Schmeidler (2003) although without the 'group' prefix. We use the prefix here to distinguish from 'simple' monotonicity defined by Kasher and Rubinstein (1997). Note that group monotonicity implies monotonicity. There is also a difference how these two papers interpret independence.

Definition 10. Let $\Gamma = (N, P, X)$ and $\Gamma' = (N, P', X)$ be two GIPs and let $i \in E$ such that $B(i) = B'(i)$. A selection rule satisfies *group independence* if $i \in f(\Gamma) \Leftrightarrow i \in f(\Gamma')$ for any such pair of problems $\Gamma, \Gamma' \in \mathcal{G}^N$. Suppose, in addition, that for all $k \neq i$ it is true that $k \in f(\Gamma) \Leftrightarrow k \in f(\Gamma')$. A selection rule satisfies *independence* if for any such pair of problems $\Gamma, \Gamma' \in \mathcal{G}^N$, $i \in f(\Gamma) \Leftrightarrow i \in f(\Gamma')$ whenever the previous condition hold.

Again the axiom introduced by Samet and Schmeidler, group independence, is stronger, i.e. it implies independence. Figure 3 features two examples which demonstrate that the TC-core violates both (group) monotonicity and (group) independence.

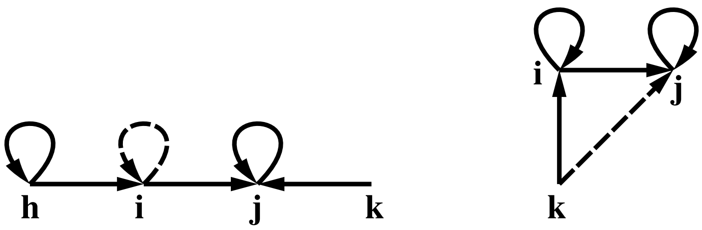

Figure 3: In both examples dashed lines represent the change in the opinion profile. (On the left) If i recommends herself then both h and i drops out from the TC-core, which is a violation of monotonicity. (On the right) If k recommends j then i drops out from the TC-core, thus independence is violated.

The final axiom of Kasher and Rubinstein's characterization is the Liberal Principle.

Definition 11. A selection rule f satisfies the *Liberal Principle* if it does not pick the empty set whenever there is an elective individual who recommends herself, and it does not pick the whole set whenever there is an elective individual who does not support herself, formally

- $\exists i \in E,\ i \in N(i) \Rightarrow f(\Gamma) \neq \emptyset$.
- $\exists i \in E,\ i \notin N(i) \Rightarrow f(\Gamma) \neq N$.

The liberal rule can be characterized in terms of strong stability. For this we introduce the liberal qualifier.

Definition 12. The *liberal qualifier* Q_L selects those individuals among the neighbours of a set of agents S who recommend themselves, formally

$$Q_L(S) = \{j \in \cup_{i \in S} N(i) \mid j \in N(j)\}.$$

Strong stability with respect to Q_L requires that each individual in the selected group recommends himself and there is a group member that testifies this (i.e. qualifies him). This second requirement is straightforward since any individual who recommends himself and is included in the group also qualifies himself. We can define an iterative algorithm in the same way as in case of the top candidate relation, which finds the *liberal core* of a game: Let each individual point to the persons among his neighbours who recommend themselves. Then the procedure is repeated with the individuals who received at least one nomination. Note that if an individual recommend himself then he also nominates himself. Thus the procedure always finishes after the second iteration. The proof of the next theorem is straightforward and is left to the reader.

Theorem 13. *A selection rule satisfies strong stability with respect to Q_L, exhaustiveness and the Liberal Principle if and only if it is the liberal rule. Moreover the liberal core coincides with the solution specified by the liberal rule.*

6 Case study – Top researchers of the nucleolus

A natural application of the above proposed model is the analysis of citation databases. Researchers acknowledge the valuable contributions of others by referencing them in their papers. Thus, journal articles can be viewed as declarations made by the authors about who is an expert of a certain subject. Naturally some of the papers are written to criticize a model, or to show the flaw in logic of some argument. The vast majority of references, however, are made to recognize the works of others.

We reviewed 88 articles of 57 authors focusing on the nucleolus and related topics (see Figure 4). Note that some of the coauthors of these papers were omitted in the analysis. Only those researchers were included in the list who are generally acclaimed to be an expert on cooperative game theory, especially on the topic of nucleolus[4].

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. Self-citations were also accounted for. The most cited author was David Schmeidler who introduced the nucleolus. This exposes a bias that is inherent in any citation analysis: older articles tend to have more citations, since references rarely point to the future[5].

Table 1 compiles the result of the citation analysis. The first column of the table shows the number of articles the researcher coauthored from the 57 articles in scope. The second column shows how many of these article contained a citation referencing the researcher. The last three column displays the well-known centrality measures: Betweenness and Closeness centrality and the PageRank value. Gold, silver and bronze colors of the cells denote the first, second and third biggest value in the column.

Betweenness centrality of a node measures the probability that a random shortest path passes through the given node. The closeness centrality of a node is the inverse of the sum of the shortest distances between the node and all other nodes reachable from it. PageRank of a node is defined recursively and depends on the number and PageRank metric of all nodes that recommend it. A node that receives many recommendations from nodes with high PageRank will have a high PageRank value itself. The formal definitions of these tree measures can be found in (Boldi and Vigna, 2014).

Table 1: Top researchers of the nucleolus. Based on the citation data of 88 articles.

Author	# of art.	# of ref.	Betweenness	Closeness	PageRank
Aarts, H.	2	16	9.736	0.0119	1.036
Arin, J.	4	16	28.724	0.0120	1.106
Aumann, R, J.	1	20	17.450	0.0114	0.908
Branzei, R.	2	9	14.551	0.0119	1.041

[4]This is admittedly a subjective element of our analysis. It is quite possible that some researchers, who deserved to be on this list, were omitted. If so, it is undoubtedly a result of oversight and not the depreciation of their work. Nevertheless, the conclusion drawn from the model is robust.

[5]Some actually do. Schmeidler introduced the nucleolus in 1969. Oddly enough, Kopelowitz already proposed a series of linear programs to compute the nucleolus in 1967.

Table 1: Top researchers of the nucleolus. Based on the citation data of 88 articles.

Author	# of art.	# of ref.	Betweenness	Closeness	PageRank
Deng, X.	2	11	12.975	0.0120	1.073
Derks, J.	3	16	44.789	0.0127	1.235
Dragan, I.	2	6	3.428	0.0099	0.536
Driessen, T. S. H.	4	24	31.795	0.0127	1.210
Elkind, E.	2	2	2.380	0.0101	0.587
Faigle, U.	4	15	18.626	0.0122	1.107
Fang, Q.	4	6	15.025	0.0122	1.107
Feltkamp, V.	3	14	21.818	0.0116	1.006
Fragnelli, V.	2	9	18.407	0.0114	0.941
Granot, D.	4	26	26.049	0.0127	1.203
Granot, F.	1	19	11.468	0.0119	1.041
Grotte, J. H.	1	6	1.089	0.0097	0.433
Hamers, H.	2	12	7.845	0.0112	0.886
Hokari, T.	2	4	3.780	0.0102	0.633
Hou, D.	2	3	0.981	0.0098	0.497
Huberman, G.	2	26	14.857	0.0122	1.103
Inarra, E.	2	8	6.647	0.0109	0.810
Jörnsten, K.	3	5	26.676	0.0112	0.927
Katsev, I. V.	4	5	18.524	0.0119	1.025
Kern, W.	6	14	15.422	0.0120	1.077
Khmelnitskaya, A. B.	2	5	9.599	0.0109	0.818
Kohlberg, E.	2	37	64.942	0.0133	1.372
Kopelowitz, A.	1	20	19.432	0.0110	0.821
Kuipers, J.	4	27	18.258	0.0127	1.194
Littlechild, S, C.	1	20	15.377	0.0111	0.877
Maschler, M.	6	50	159.629	0.0169	1.888
Megiddo, N.	3	29	35.683	0.0125	1.187
Montero, M.	3	2	0.941	0.0100	0.557
Núñez, M.	3	2	3.416	0.0103	0.644
Okamoto, Y.	1	3	1.825	0.0102	0.624
Orshan, G.	2	6	1.211	0.0104	0.668
Owen, G.	2	39	68.788	0.0137	1.422
Paulusma, D.	3	5	6.049	0.0111	0.857
Peleg, B.	2	49	135.792	0.0159	1.759
Peters, H.	2	2	13.178	0.0110	0.815
Potters, J.	8	32	105.505	0.0152	1.655
Raghavan, T. E. S.	3	25	38.708	0.0132	1.305
Reijnierse, H.	4	25	41.420	0.0132	1.308
Sankaran, J. K.	1	8	1.451	0.0102	0.579
Schmeidler, D.	1	55	208.239	0.0175	1.942
Serrano, R.	2	5	4.184	0.0103	0.661

Table 1: Top researchers of the nucleolus. Based on the citation data of 88 articles.

Author	# of art.	# of ref.	Betweenness	Closeness	PageRank
Shapley, L, S.	1	50	148.548	0.0164	1.821
Snijders, C.	1	5	32.349	0.0132	1.298
Sobolev, A, I.	1	27	0.228	0.0097	0.423
Solymosi, T.	6	26	36.310	0.0119	1.087
Sudhölter, P.	6	18	29.035	0.0118	1.050
Tijs, S.	6	31	97.916	0.0149	1.614
van den Brink, R.	2	3	5.635	0.0110	0.828
Wallmeier, E,	1	10	3.693	0.0101	0.593
Yanovskaya, E.	2	3	6.477	0.0108	0.775
Zarzuelo, J. M.	2	6	6.612	0.0103	0.627
Zhou, L.	1	5	1.347	0.0098	0.464
Zhu, W. R.	2	15	5.179	0.0115	0.941

The TC-algorithm selects Michael Maschler and David Schmeidler as the core experts. This conclusion is backed up by the three classic centrality measure which rank these two game theorist as the top 2 experts among the researchers of this field. The eliminative TC-algorithm also selects Maschler and Schmeidler but also includes Elon Kohlberg, Bezalel Peleg and Lloyd S. Shapley.

Indeed all these researchers had a great impact in the development of the nucleolus. Schmeidler, as we mentioned earlier, introduced the solution concept itself. 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". They not only gave an intuitive geometric description of the nucleolus but also devised a linear programming framework to compute it. This was the first LP that could actually be implemented to determine the nucleolus. All the previous attempts needed way too many constraints (one in particular worked with $\mathcal{O}(2^n!)$) or had other weaknesses. Finally Kohlberg introduced a criterion, which can be used to verify whether an allocation is the nucleolus or not. The Kohlberg-criterion has great practical and theoretical significance.

All centrality measures proved to be an appropriate tool in identifying the key researchers of this field. The two algorithms we proposed in this paper performed no worse. Both the TC- and eliminative TC-algorithms picked the top choices of these measures. This indicates that the TC-algorithms can be effectively used to create a shortlist of experts.

7 Conclusion and future research

We extended the group identification framework in two ways. First we changed the focus from subjective social categories like ethnicity or religion to groups whose formation encompasses both subjective and objective elements. Secondly we lifted the requirements that selection rules must satisfy monotonicity and independence. We have shown that there are interesting groups and selection rules that identify such groups, which are not compatible with either monotonicity or independence.

The Top Candidate and eliminative Top Candidate algorithms are effective complementary tools that can be applied in citation analysis. By narrowing down the possible group members we can determine the core experts in the community. The Top Candidate algorithm is axiomatically well-founded - strong stability and exhaustiveness are natural axioms, which – with the appropriate qualifier – also characterize the liberal rule. It remains an open question how the eliminative Top Candidate algorithm can be characterized.

The new characterization of the liberal rule suggests that there is a connection between the axiomatic approach used in this paper and the ideas of Miller (2008) and Samet and Schmeidler (2003). In particular it is an interesting question how strong stability relates to affirmative and exclusive self-determination, or the meet and join separability axioms.

Apart from a recent working paper by Cho and Saporiti (2015), which studies strategy-proofness in a group identification setting, very little has been done concerning the incentives of the individuals. Here we did not delve into the strategic aspect of the model. One could argue that the Top Candidate algorithm is somewhat vulnerable to strategic manipulation. A possible scenario is when an individual cites only his own papers and works that belong to lesser known researchers. Such an individual forms a top candidate component by himself, hence he is selected by the Top Candidate algorithm. This is quite natural for researchers with exceptional skills, however even a mediocre individual can trick himself into the selection if he consistently avoids to cite the prominent figures of his field.

Finally it is worth to note that the literature already turned toward the analysis of social networks with multiple groups. The popularity of network services like Facebook or Twitter suggests that there will be an increasing demand for tools that can identify key players in networks (e.g. for advertising purposes). Chen et al. (2010) examined the problem from game theoretical point of view, while Nicolas (2007) used the group identification framework to study such models. A straightforward question is how to extend the Top Candidate algorithm to handle multiple groups and how to solve the arising algorithmic and axiomatic difficulties.

References

Amatriain, X., Lathia, N., Pujol, J. M., Kwak, H., and Oliver, N. (2009). The wisdom of the few: A collaborative filtering approach based on expert opinions from the web. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 532–539.

Boldi, P. and Vigna, S. (2014). Axioms for centrality. *Internet Mathematics*, 10(3-4):222–262.

Carchiolo, V., Longheu, A., M., M., and Mangioni, M. (2015). Searching for experts in a context-aware recommendation network. *Computers in Human Behavior*.

Çengelci, M. and Sanver, M. (2010). Simple Collective Identity Functions. *Theory and Decision*, 68(4):417–443.

Chen, W., Liu, Z., Sun, X., and Wang, Y. (2010). A game-theoretic framework to identify overlapping communities in social networks. *Data Mining and Knowledge Discovery*, 21(2):224–240.

Cho, W. J. and Saporiti, A. (2015). Incentives, fairness, and efficiency in group identification. Discussion Paper Series EDP-1501, University of Manchaster.

Dimitrov, D., Sung, S. C., and Xu, Y. (2007). Procedural group identification. *Mathematical Social sciences*, 54:137–146.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1):131–152.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 102(46), pages 16569–16572.

John, A. and Seligmann, D. (2006). Collaborative tagging and expertise in the enterprise. In *Proceedings of the WWW 2006 Collaborative Web Tagging Workshop*, Edinburgh, Scotland.

Ju, B.-G. (2010). Individual powers and social consent: an axiomatic approach. *Social Choice and Welfare*, 34(4):571–596.

Kasher, A. and Rubinstein, A. (1997). On the question "who is a j?": a social choice approach. *Logique et Analyse*, 160:385–395.

Laslier, J.-F. (1997). *Tournament Solutions and Majority Voting*. Studies in Economic Theory (Book 7). Springer.

Miller, A. (2008). Group identification. *Games and Economic Behavior*, 63:188–202.

Nicolas, H. (2007). "i want to be a j!": Liberalism in group identification problems. *Mathematical Social Sciences*, 54:59–70.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web.

Pal, A. and Konstan, J. A. (2010). Expert identification in community question answering: Exploring question selection bias. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pages 1505–1508.

Samet, D. and Schmeidler, D. (2003). Between liberalism and democracy. *Journal of Economic Theory*, 110:213–233.

Sung, S. C. and Dimitrov, D. (2005). On the axiomatic characterization of "who is a j?". *Logique et Analyse*, 189-192:101–112.

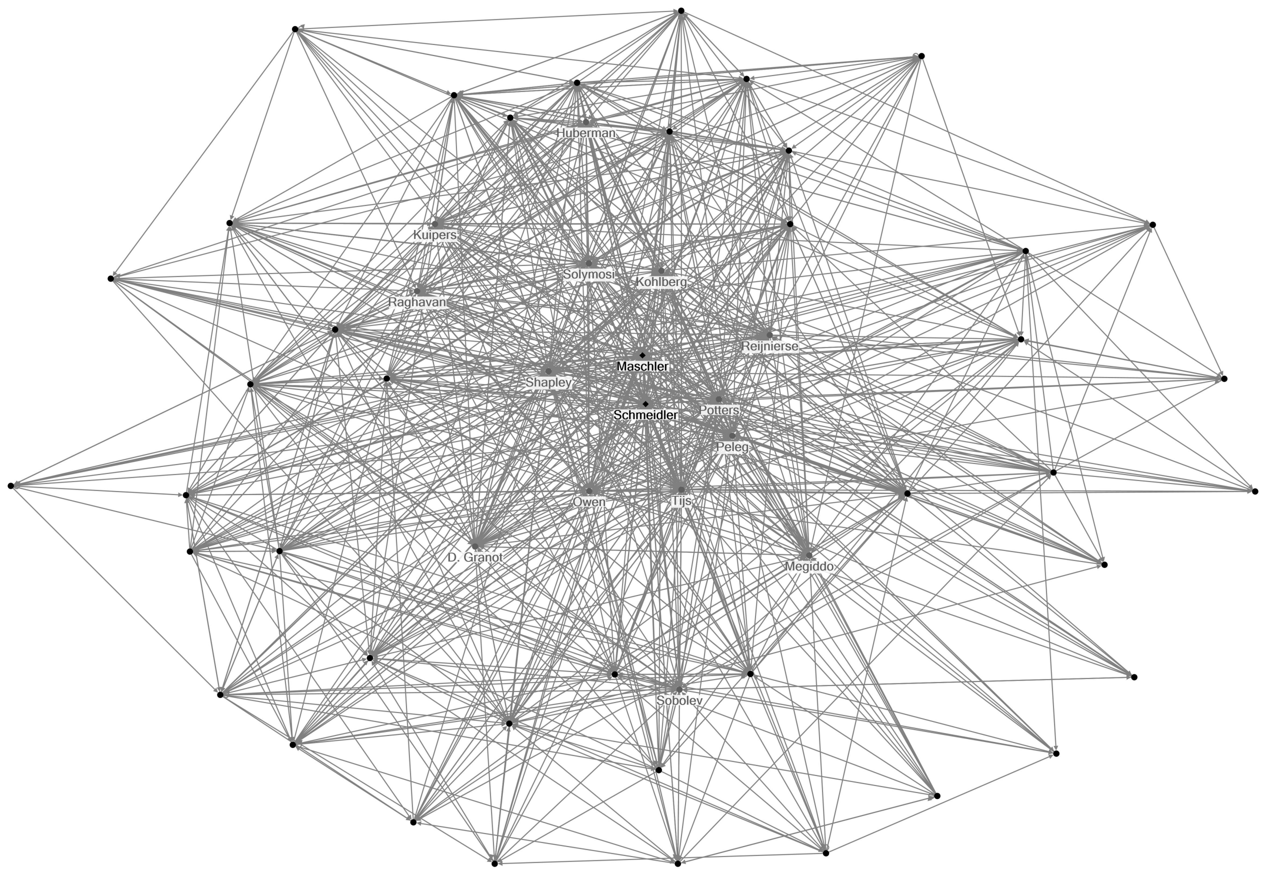

Figure 4: Citation data of 88 articles focusing on the nucleolus and related topics. The opinion graph has 57 nodes (authors) and 937 arcs (references). Author nodes with at least 25 references are labeled. Visualized by NodeXL using Harel-Koren Fast Multiscale method. The Top Candidate algorithm selects Michael Maschler and David Schmeidler as the core experts.