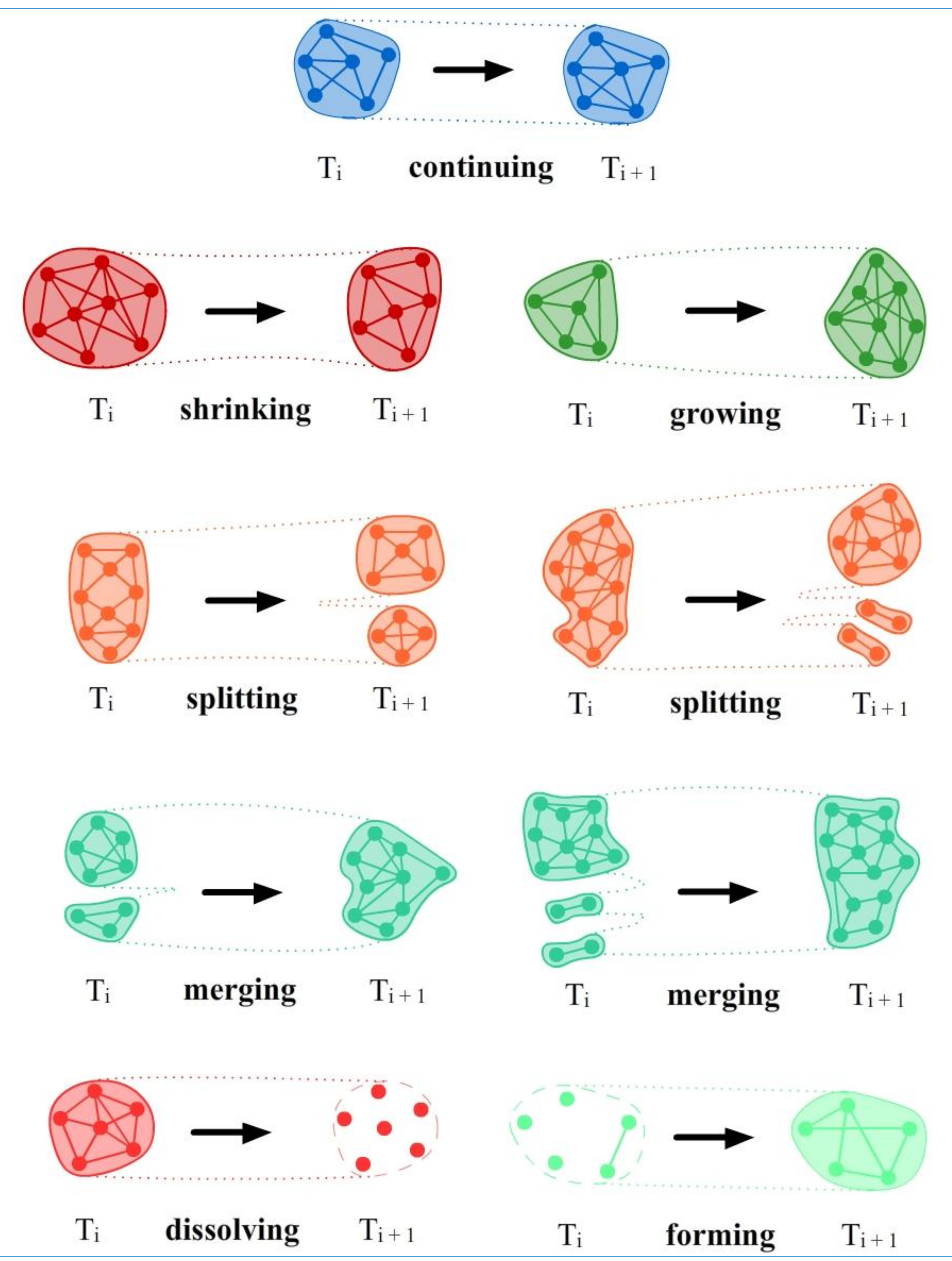
GROUP EVOLUTION DISCOVERY

Piotr Bródka, Stanisław Saganowski, Przemysław Kazienko
piotr.brodka@pwr.wroc.pl, stanislaw.saganowski@pwr.wroc.pl, kazienko@pwr.wroc.pl

Wrocław University of Technology, Wyb.Wyspiańskiego 27, 50-370 Wrocław, Poland

Group extraction and their evolution are among the topics which arouse the greatest interest in the SNA domain. However, while the group extraction methods in social networks are developed very dynamically, the methods of **group evolution discovery are still '*uncharted territory*'**. In recent years, only few methods for tracking changes of social groups have been proposed [2,3,5,6]. Therefore we present the new method for the group evolution discovery called ***GED***.

1. Temporal Social Network and Events in Group Evolution

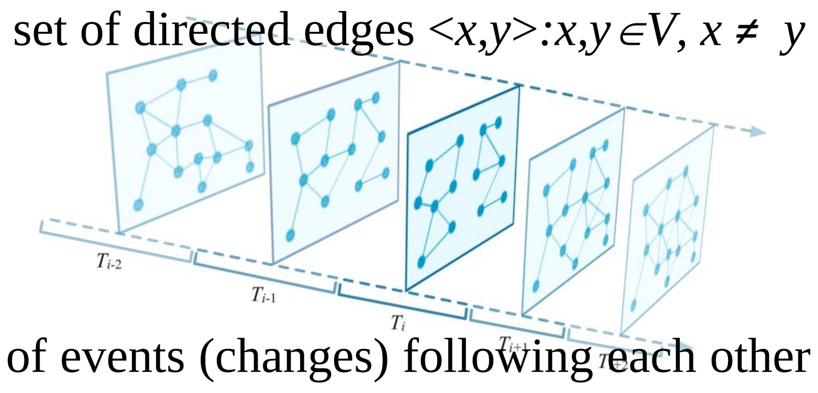


Temporal social network *TSN:* is a list of succeeding timeframes (time windows) *T*. Each timeframe is in fact one social network *SN*(*V*,*E*) where *V* – is a set of vertices and *E* is a set of directed edges $<x,y>: x,y \in V, x \neq y$

$$TSN = < T_1, T_2, \ldots, T_m >, \quad m \in N$$

$$T_i = SN_i(V_i, E_i), \quad i = 1,2,\ldots,m$$

$$E_i = < x, y >: x, y \in V_i, x \neq y \quad i = 1,2,\ldots,m$$



Evolution of a social community can be represented by a sequence of events (changes) following each other in the successive timeframes within the *TSN*. Possible **events in social group evolution** are:

1. ***Continuing* (stagnation)** – two following groups are identical or differ only a little (size remain the same).
2. ***Shrinking*** – some nodes have left the group, making its size smaller than in the previous time window.
3. ***Growing*** – some new nodes have joined the group, making its size bigger than in the previous time window.
4. ***Splitting*** – the group splits into 2 or more groups in the next time window when few groups from T_{i+1} consist of members of one group from T_i. Two types of splitting: (1) *equal* - the contribution of the groups in The split group is almost the same and (2) *unequal* - one of the groups has much greater contribution in the split group, which for this one group the event might be similar to shrinking.
5. ***Merging*** – merging several other groups when one group from T_{i+1} consist of two or more groups from the previous time T_i. Merge might be (1) *equal* - the contribution of the groups in the merged group is almost the same, or (2) unequal - one of the groups has much greater contribution into the merged group (for the biggest group the merging might be similar to growing).
6. ***Dissolving* –** a group ends its life and does not occur in the next time window.
7. ***Forming*** – a group which has not existed in the previous time T_i appears in T_{i+1}. A group can be inactive over several timeframes – it is treated as dissolving of the first group and forming again of the, second, new one.

2. The Inclusion Measure

Key component of *GED* - a **new measure** called ***inclusion***. It allows to evaluate the inclusion of one group in another:

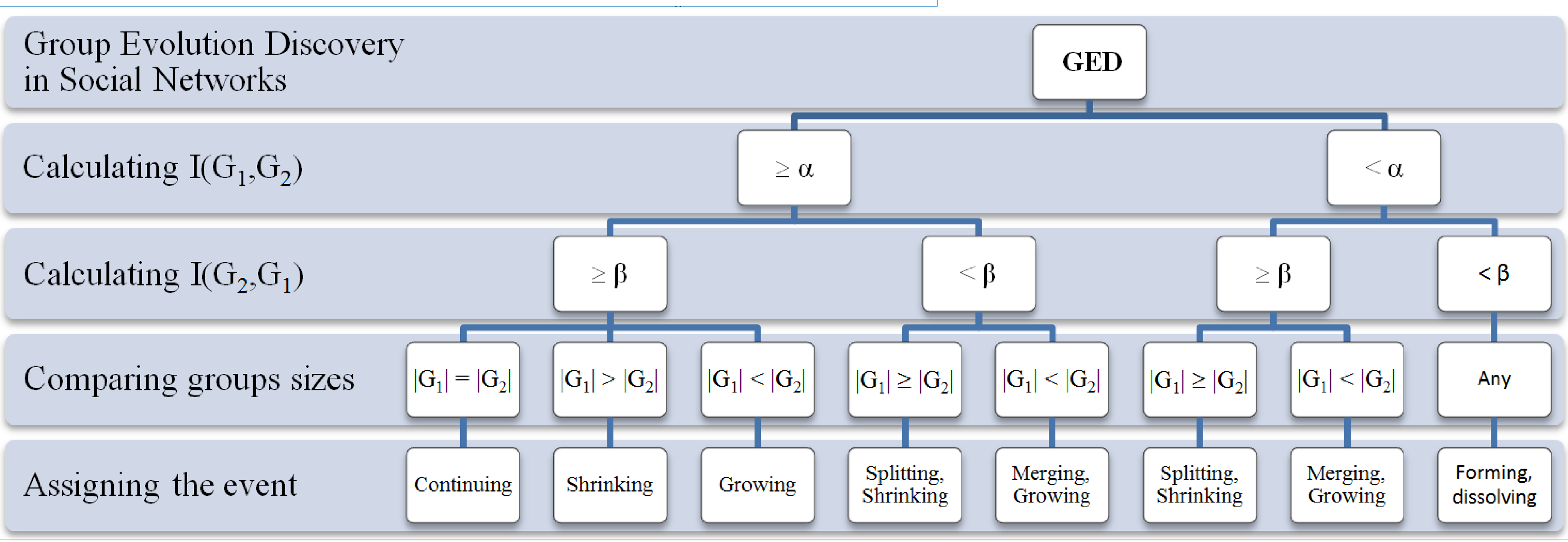
$$I(G_1, G_2) = \overbrace{\frac{|G_1 \cap G_2|}{|G_1|}}^{\text{group quantity}} \cdot \underbrace{\frac{\sum_{x \in (G_1 \cap G_2)} SP_{G_1}(x)}{\sum_{x \in (G_1)} SP_{G_1}(x)}}_{\text{group quality}}$$

The *GED* method, to match two groups from consecutive timeframes takes into consideration both, the **quantity and quality** of the group members. The **quantity** is reflected by the first part of the ***inclusion*** measure, i.e. what portion of G_1 members is shared by G_2, whereas the **quality** is expressed by the second part of the *inclusion* measure, namely what contribution of important members of G_1 is shared by G_2. It provides a balance between the groups, which contain many of the less important members and groups with only few but key members. To indicate user importance one of the centrality measures may be used. For this presentation we have utilized SP measure [4].

3. GED – Group Evolution Discovery Method

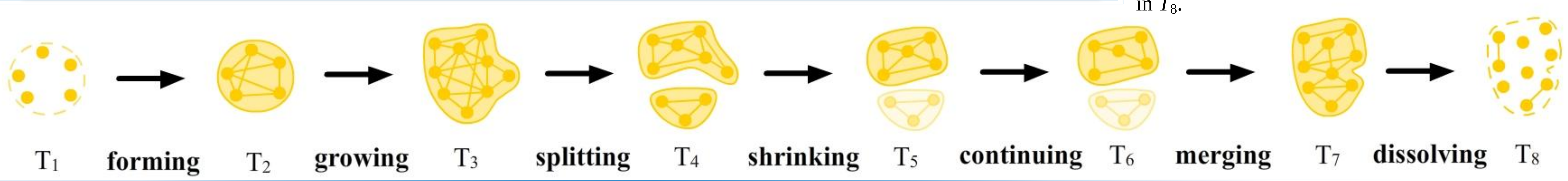
Input: Groups in *TSN* are extracted by **any community detection** algorithm for each timeframe T_i . Calculated any **user importance measure**.

1. For each pair $<G_1, G_2>$ in timeframes T_i and T_{i+1} inclusion of G_1 in G_2 and G_2 in G_1 is computed.
2. Based on **inclusion and size** of two groups one type of event may be assigned:
   a. ***Continuing***: $I(G_1,G_2) > \alpha$ and $I(G_2,G_1) > \beta$ and $|G_1| = |G_2|$
   b. ***Shrinking***: $I(G_1,G_2) > \alpha$ and $I(G_2,G_1) > \beta$ and $|G_1| > |G_2|$ OR $I(G_1,G_2) < \alpha$ and $I(G_2,G_1) > \beta$ and $|G_1| > |G_2|$ and there is only one matching event between G_2 and all groups in T_i
   c. ***Growing***: $I(G_1,G_2) > \alpha$ and $I(G_2,G_1) > \beta$ and $|G_1|<|G_2|$ OR $I(G_1,G_2) > \alpha$ and $I(G_2,G_1) < \beta$ and $|G_1| > |G_2|$ and there is only one matching event between G_1 and all groups in the next time window T_{i+1}
   d. ***Splitting***: $I(G_1,G_2) < \alpha$ and $I(G_2,G_1) > \beta$ and $|G_1| > |G_2|$ and there is more than one match (matching events) between G_2 and all groups in the previous time window T_i
   e. ***Merging***: $I(G_1,G_2) > \alpha$ and $I(G_2,G_1) < \beta$ and $|G_1| > |G_2|$ and there is more than one match (matching events) between G_1 and all groups in the next time window T_{i+1}
   f. ***Dissolving***: for G_1 in T_i and each group G_2 in T_{i+1} $I(G_1,G_2) < 10\%$ and $I(G_2,G_1) < 10\%$
   g. ***Forming***: for G_2 in T_{i+1} and each group G_1 in T_i $I(G_1,G_2) < 10\%$ and $I(G_2,G_1) < 10\%$



The indicators α and β are **GED parameters** and are used to **adjust the method** to particular SN and community detection method. Based on experiments [1] authors suggest the values of α and β be from range **[50%;100%]**

After running *GED* for all timeframes **we can extract group evolution** (see below). The group forms in T_2, then by gaining new nodes grows in T_3, splits into two groups in T_4, then by losing one node the bigger group is shrinking in T_5, both groups continue over T_6, next they merge with the third group in T_7, and finally the group dissolves in T_8.



1 Bródka P., Saganowski S., Kazienko P.: Group Evolution Discovery in Social Networks. ASONAM 2011, Taiwan, 25-27 July 2011, IEEE Computer Society, 2011, pp 247-253
2 Chakrabarti D, Kumar R, Tomkins A, Evolutionary Clustering, KDD'06, August 20–23, 2006, Philadelphia, Pennsylvania, USA.
3 Kim, M.-S., and J. Han, 2009, in Proceedings of 2009 Int. Conf. on Very Large Data Bases. A Particle and Density Based Evolutionary Clustering Method for Dynamic Networks
4 Musial K., Kazienko K., Bródka P. User position measures in social networks. SNA-KDD'09. ACM, New York, NY, USA, Article 6 , 9 pages.
5 Palla, G., Barabási, A.L., and Vicsek, T. Quantifying social group evolution. Nature 446, (2007), 664-667.
6 Sun J., Papadimitriou S., Yu P., Faloutsos C, GraphScope: Parameter-free Mining of Large Time-evolving Graphs

Acknowledgments.

The work was supported by: Fellowship co-Financed by the European Union within the European Social Fund, The Polish Ministry of Science and Higher Education, the research project, 2010-13, The training in "Green Transfer" co-financed by the EU from the European Social Fund