Wrocław University of Technology
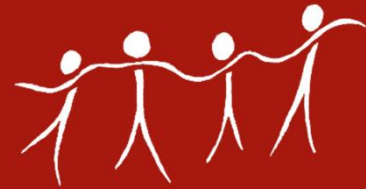
Group Evolution Discovery in Social Networks

Piotr Bródka, Stanisław Saganowski, Przemysław Kazienko

Wrocław University of Technology, Poland

Research Engineering Center, Poland

piotr.brodka@pwr.wroc.pl,
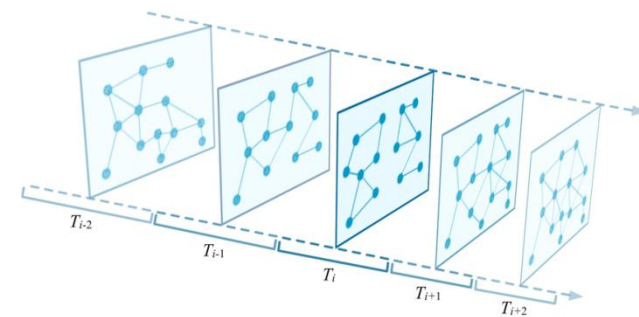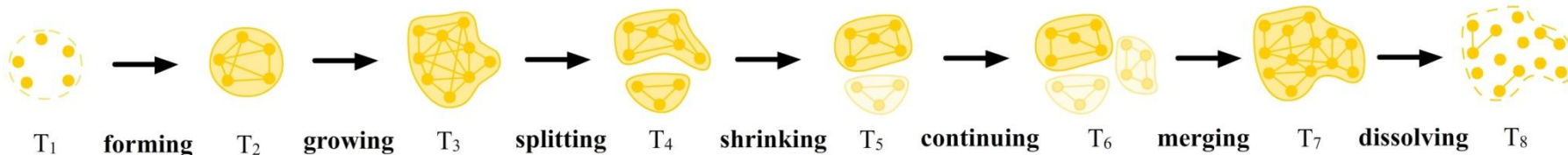stanislaw.saganowski@pwr.wroc.pl, kazienko@pwr.wroc.pl

Agenda

✓ Problem description and motivation

✓ Basic concepts

✓ Group evolution

✓ Inclusion measure

✓ GED

✓ Experiments

✓ Final remarks

Tracking Group Evolution in Social Networks

- **Groups extraction** is nice

- ... but **group evolution prediction** is nicer ...

- ... so we need to **identify changes** in group evolution.

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

| T_1 | forming | T_2 | growing | T_3 | splitting | T_4 | shrinking | T_5 | continuing | T_6 | merging | T_7 | dissolving | T_8 |

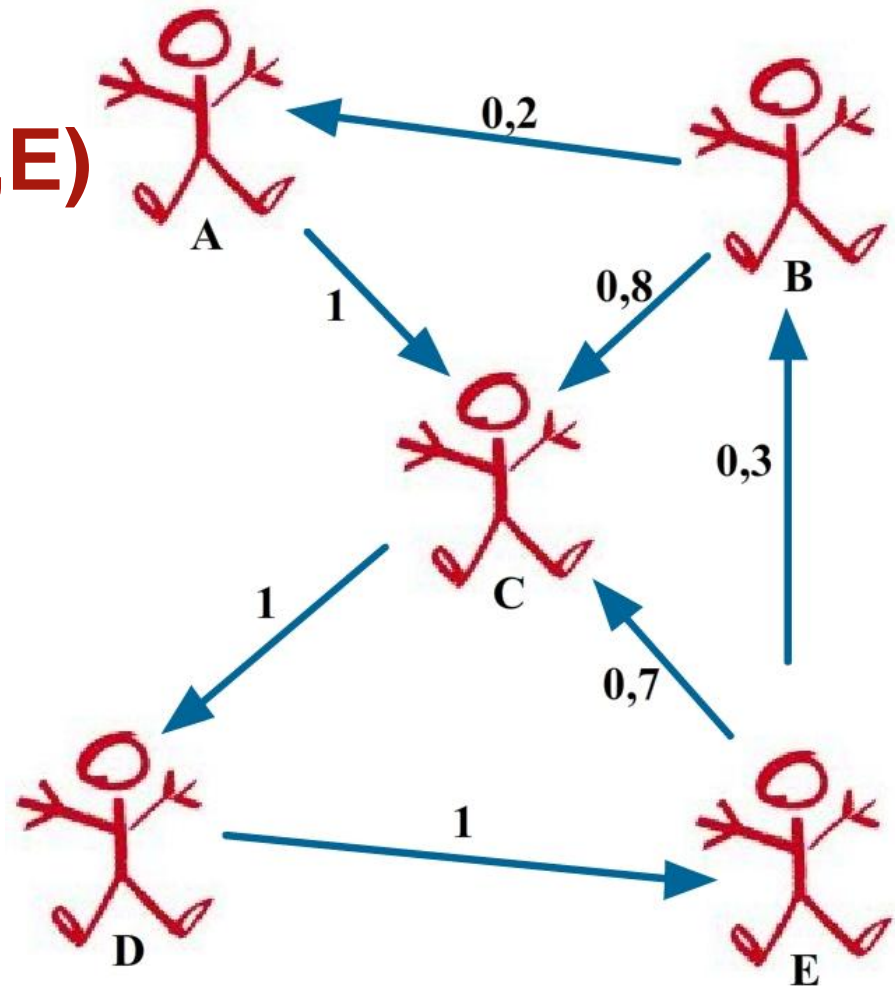# Basic Concepts: Social Network

Social network: **SN(V,E)**

V – a set of **vertices**

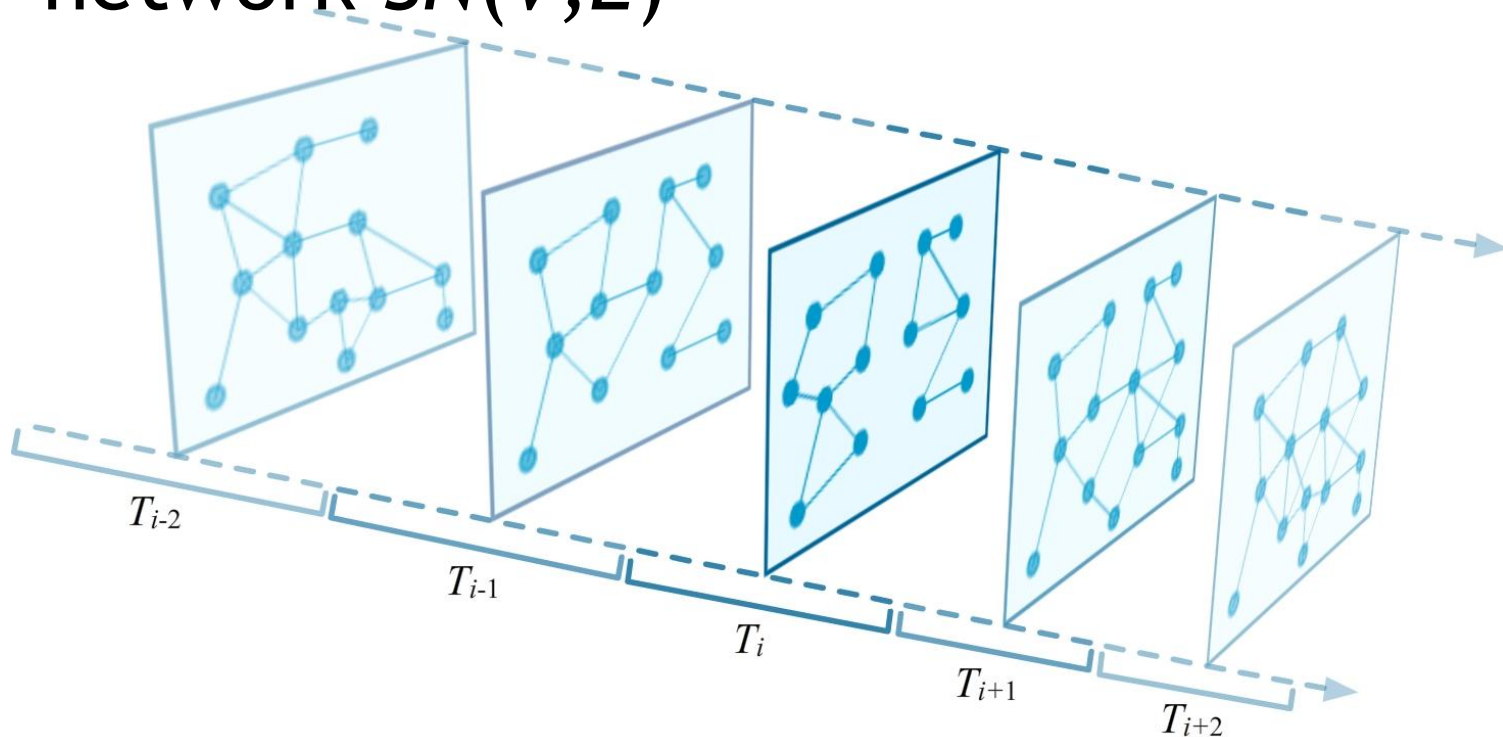*E* – a set of directed **edges** *<x,y>:x,y∈V*

Basic Concepts: Temporal Social Network

Social Network Group
@ Wroclaw University of Technology

TSN - a list of following timeframes (time windows) *T, e*ach is a social network *SN(V,E)*

T_{i-2}

T_{i-1}

T_i

T_{i+1}

T_{i+2}

Basic Concepts:
Group (Community)

Social Network Group
@ Wroclaw University of Technology

- **No** commonly accepted group **definition**

- A group is a set of people, who have **strong mutual** (internal) relationships and weak with people outside the group (external)

- **Group G** in the social network $SN(V,E)$ is a **subset of vertices** ($G \subseteq V$), extracted using any method (clustering algorithm)

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN

  - **Continuing**
  - Shrinking
  - Growing
  - Splitting
  - Merging
  - Dissolving
  - Forming

T_i **continuing** T_{i+1}

Group Evolution

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN
  - Continuing
  - **Shrinking**
  - Growing
  - Splitting
  - Merging
  - Dissolving
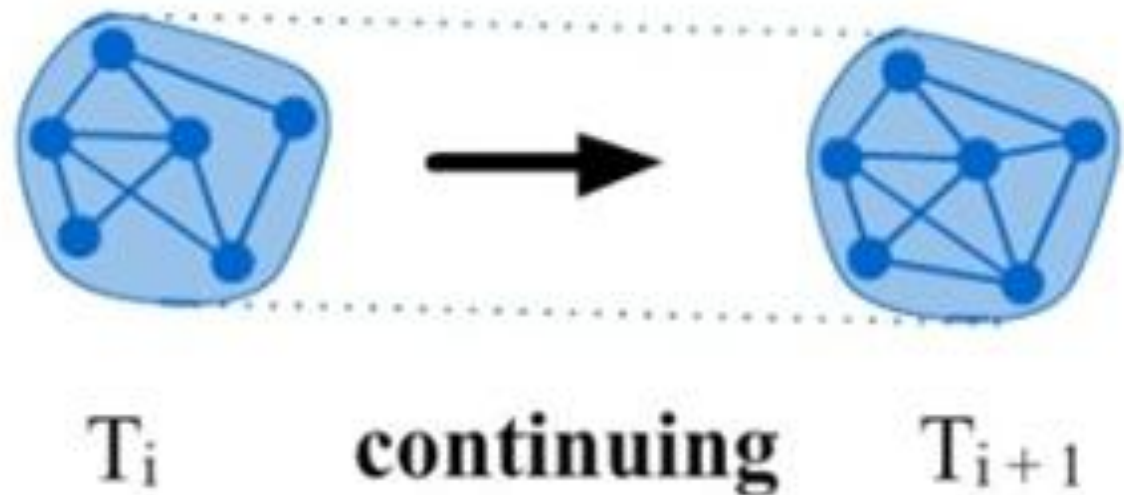 - Forming

T_i shrinking T_{i+1}

Group Evolution

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN
  - Continuing
  - Shrinking
  - **Growing**
  - Splitting
  - Merging
  - Dissolving
  - Forming

T_i growing T_{i+1}

Group Evolution

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN
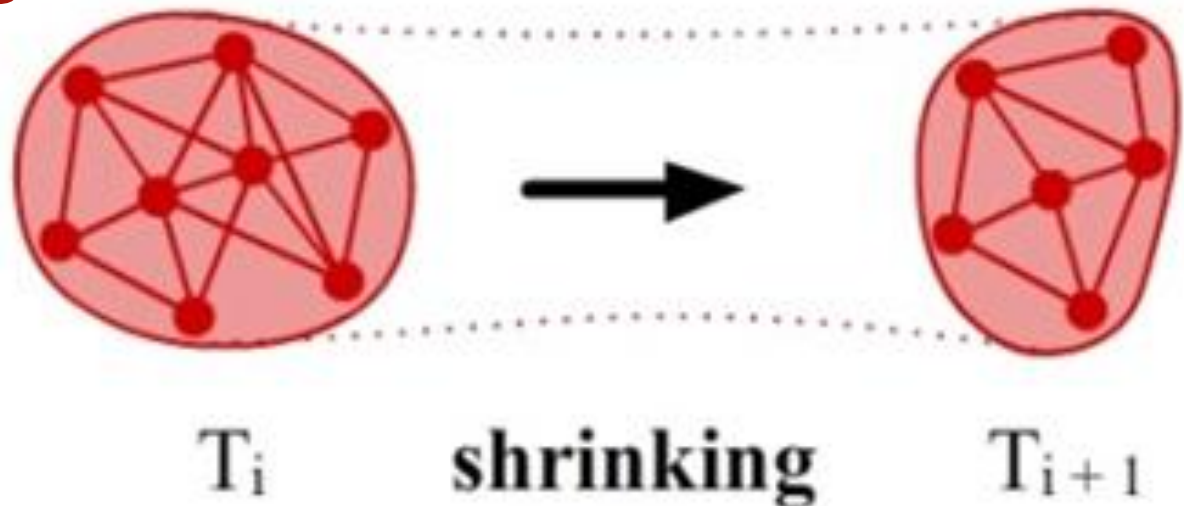 - Continuing
  - Shrinking
  - Growing
  - **Splitting**
  - Merging
  - Dissolving
  - Forming

T_i **splitting** T_{i+1}

Group Evolution

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN
  - Continuing
  - Shrinking
  - Growing
  - Splitting
  - **Merging**
  - Dissolving
  - Forming

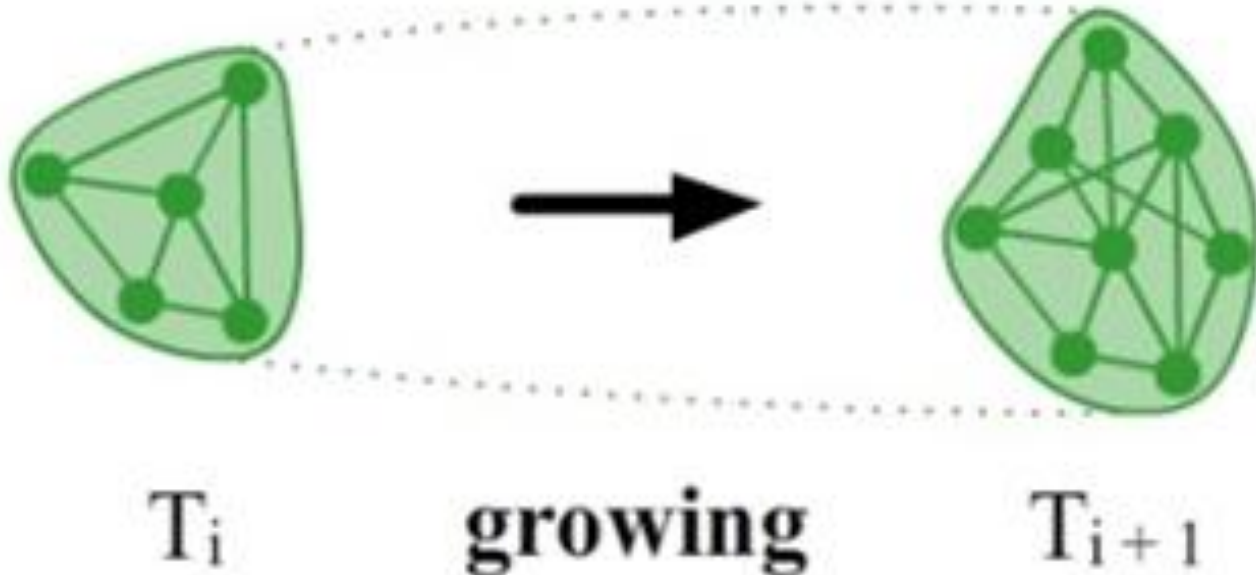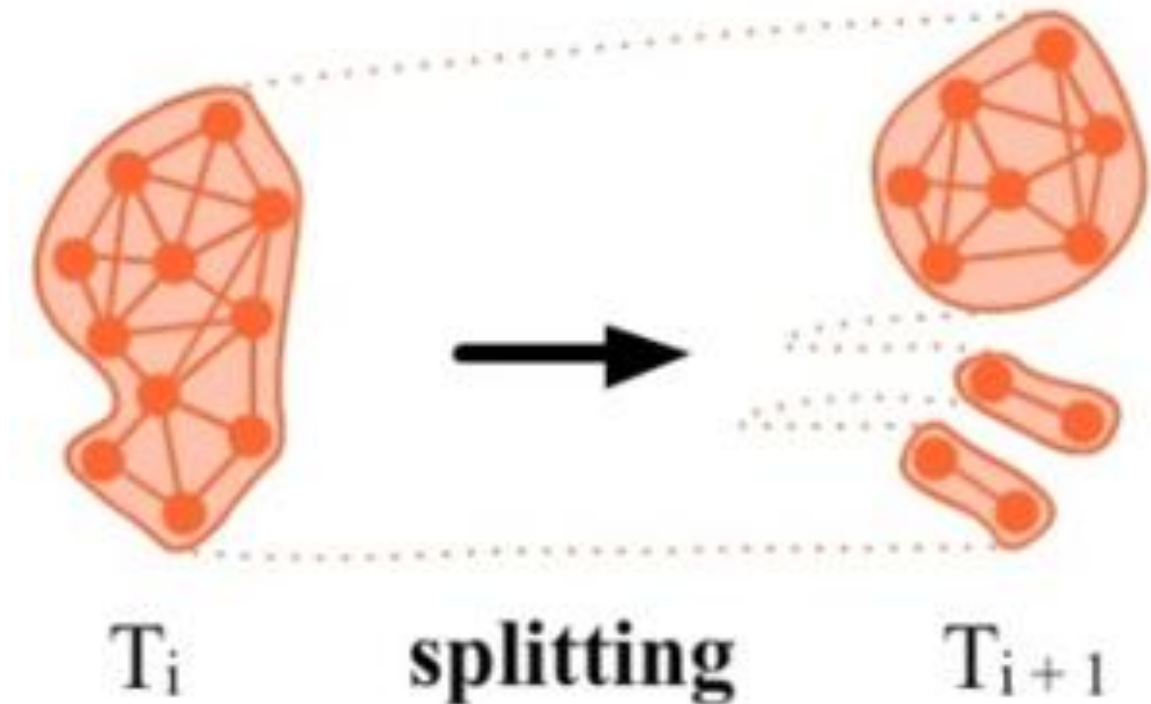T_i merging T_{i+1}

Group Evolution

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN
  - Continuing
  - Shrinking
  - Growing
  - Splitting
  - Merging
  - **Dissolving**
  - Forming

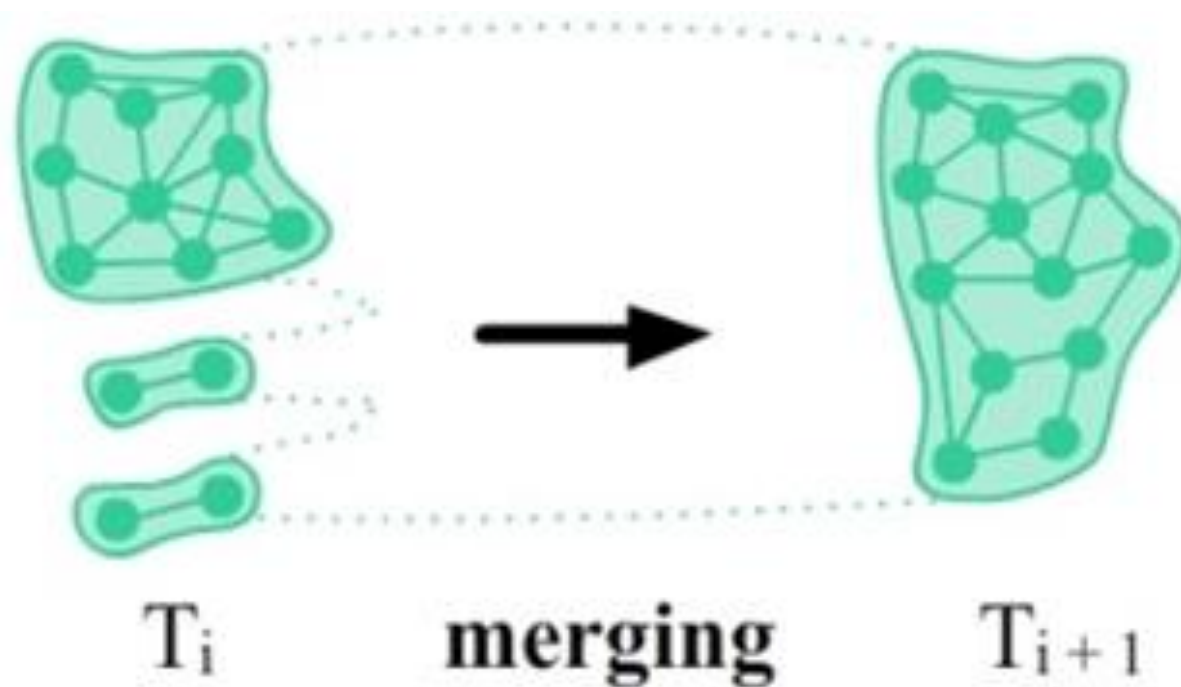T_i dissolving T_{i+1}

Group Evolution

- Group evolution is a **sequence of events** succeeding each other in the successive time windows within TSN
  - Continuing
  - Shrinking
  - Growing
  - Splitting
  - Merging
  - Dissolving
  - **Forming**

T_i forming T_{i+1}
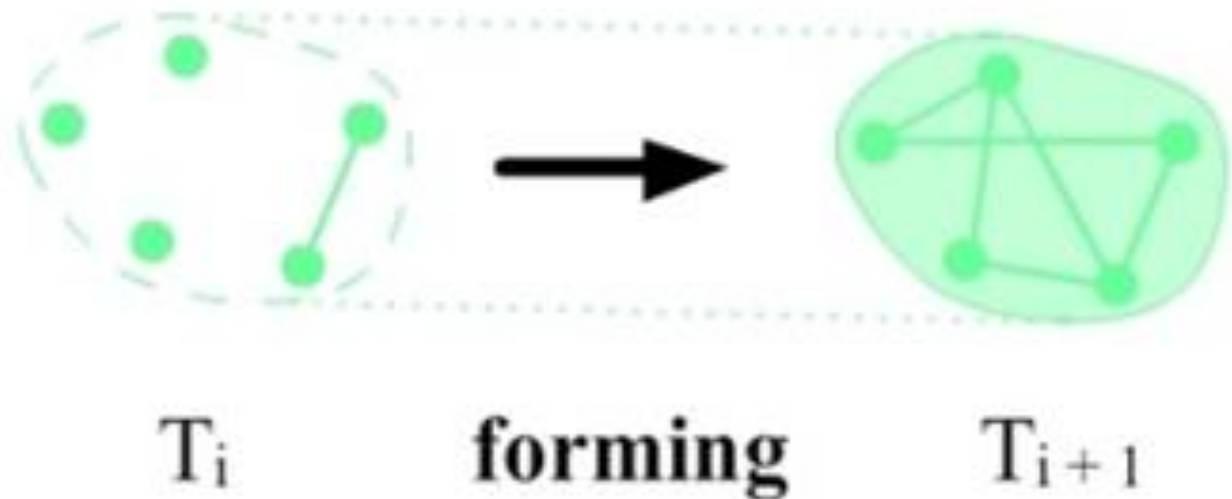
GED Method: Introduction

- *GED* (Group Evolution Discovery) method takes into account

  - **quantity** of the group members

  - **quality** of the group members

- Members quality: any **centrality measure**

  - social position and degree centrality measures was utilized in the experiments

GED method: Inclusion measure

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

Group **quantity**

Group **quality**

GED – Group Evolution Discovery Method

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

For each pair of groups <G_1, G_2> in consecutive timeframes T_i and T_{i+1} inclusion of G_1 in G_2 and G_2 in G_1 is counted according to equations (3).

Based on inclusion and size of two groups one type of event may be assigned:

> *Continuing*: $I(G_1,G_2) \geq a$ and $I(G_2,G_1) \geq B$ and $|G_1| = |G_2|$
>
> *Shrinking*: $I(G_1,G_2) \geq a$ and $I(G_2,G_1) \geq B$ and $|G_1| > |G_2|$ OR $I(G_1,G_2) < a$ and $I(G_2,G_1) \geq B$ and $|G_1| \geq |G_2|$ and there is only one match (matching event) between G_2 and all groups in the previous time window T_i
>
> *Growing*: $I(G_1,G_2) \geq a$ and $I(G_2,G_1) \geq B$ and $|G_1| < |G_2|$ OR $I(G_1,G_2) \geq a$ and $I(G_2,G_1) < B$ and $|G_1| \leq |G_2|$ and there is only one match (matching event) between G_1 and all groups in the next time window T_{i+1}
>
> *Splitting*: $I(G_1,G_2) < a$ and $I(G_2,G_1) \geq B$ and $|G_1| \geq |G_2|$ and there is more than one match (matching events) between G_2 and all groups in the previous time window T_i
>
> *Merging*: $I(G_1,G_2) \geq a$ and $I(G_2,G_1) < B$ and $|G_1| \leq |G_2|$ and there is more than one match (matching events) between G_1 and all groups in the next time window T_{i+1}
>
> *Dissolving*: for G_1 in T_i and each group G_2 in T_{i+1} $I(G_1,G_2) < 10\%$ and $I(G_2,G_1) < 10\%$
>
> *Forming*: for G_2 in T_{i+1} and each group G_1 in T_i $I(G_1,G_2) < 10\%$ and $I(G_2,G_1) < 10\%$

GED – Group Evolution Discovery Method

Social Network Group
@ Wroclaw University of Technology

Input: *TSN* in which at each timeframe T_i groups are extracted by any community detection

Group Evolution Discovery in Social Networks	GED

Calculating $I(G_1,G_2)$: $\geq \alpha$ / $< \alpha$

Calculating $I(G_2,G_1)$: $\geq \beta$ / $< \beta$ / $\geq \beta$ / $< \beta$

Comparing groups sizes: $|G_1| = |G_2|$ / $|G_1| > |G_2|$ / $|G_1| < |G_2|$ / $|G_1| \geq |G_2|$ / $|G_1| < |G_2|$ / $|G_1| \geq |G_2|$ / $|G_1| < |G_2|$ / Any

Assigning the event: Continuing / Shrinking / Growing / Splitting, Shrinking / Merging, Growing / Splitting, Shrinking / Merging, Growing / Forming, dissolving

Merging: $I(G_1,G_2) \geq a$ and $I(G_2,G_1) < B$ and $|G_1| \leq |G_2|$ and there is more than one match (matching events) between G_1 and all groups in the next time window T_{i+1}

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%
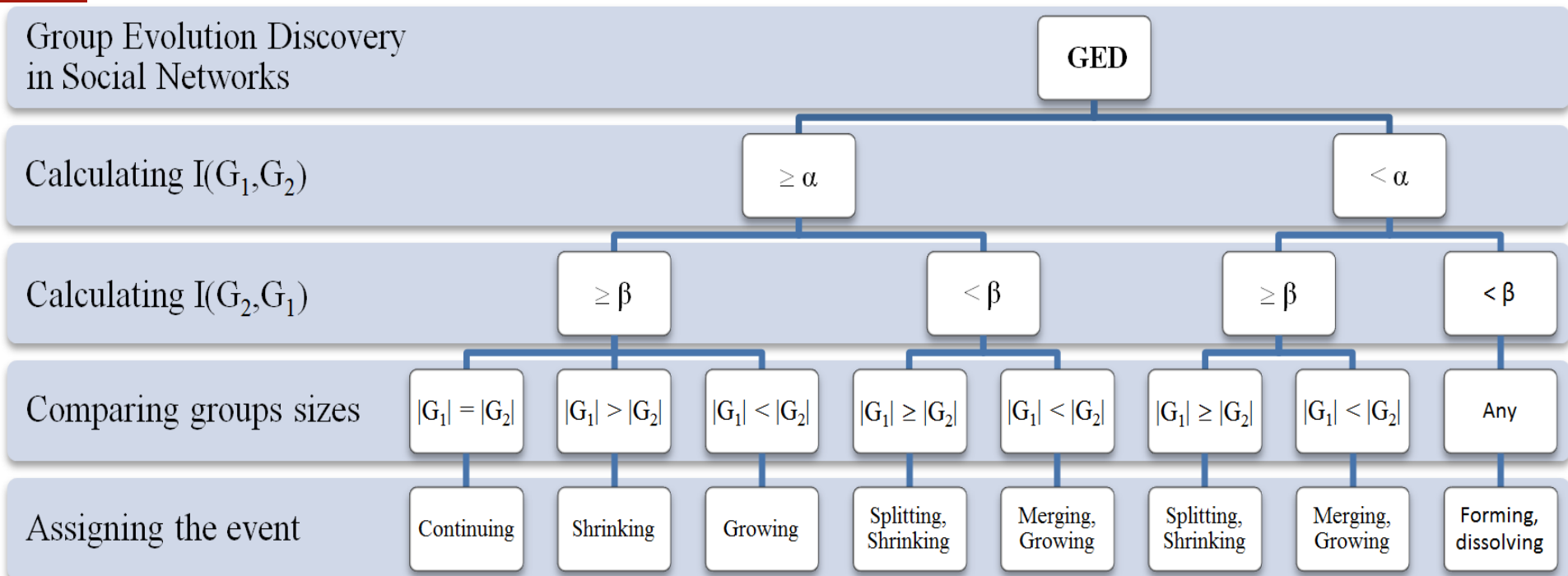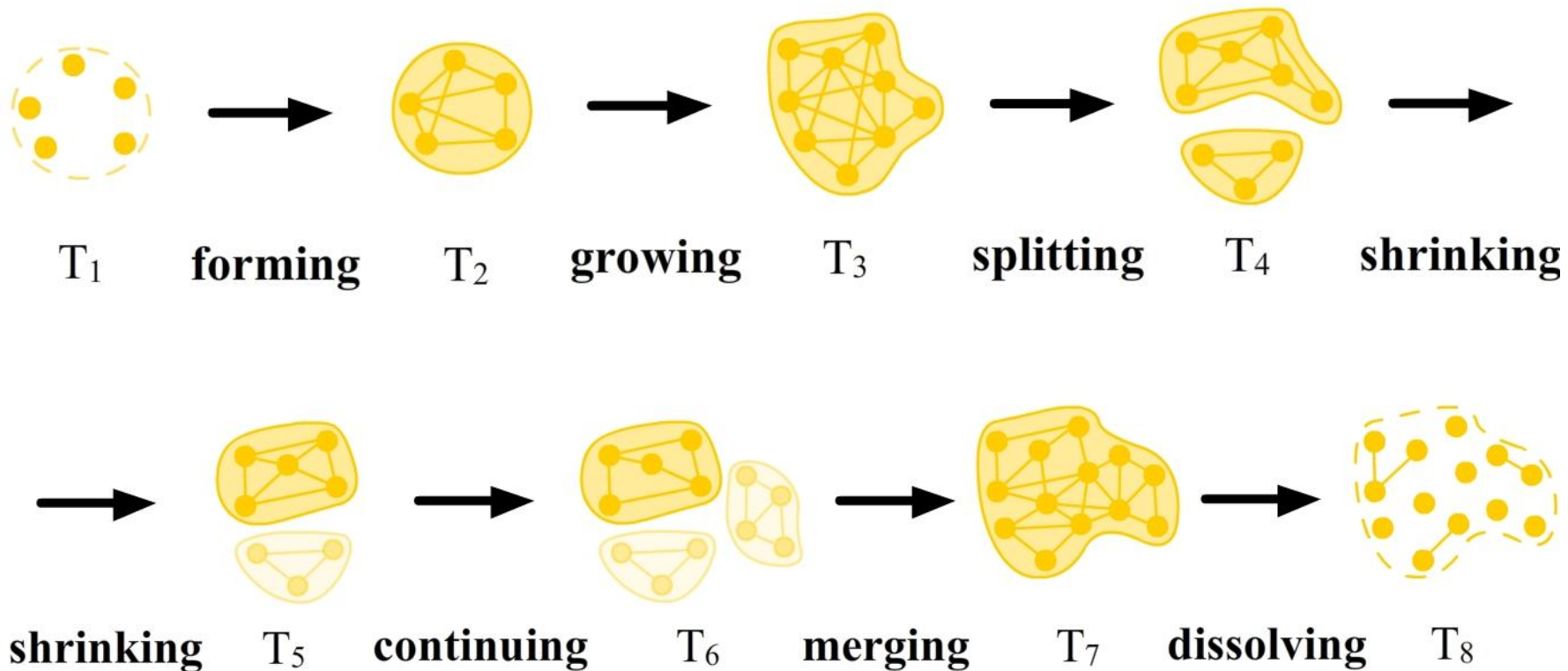
Forming: for G_2 in T_{i+1} and each group G_1 in T_i $I(G_1,G_2)$ < 10% and $I(G_2,G_1)$ < 10%

GED: Following Events

Wrocław University of Technology

Experiments: Setup

Social Network Group
@ Wroclaw University of Technology

- Data Set
  - Staff email exchange from WrUT (**270K**+ emails, **2 years**)
  - **5,845** nodes and **149,344** edges
  - Fourteen moving **90-days frames** (overlap 45 days)
- Community extraction methods
  - Fast modularity optimization (disjoint groups)
  - CPM (overlapping groups)
- Methods for tracking group evolution
  - by Asur et al.
  - by Palla et al.
  - **GED**

Experiments: Results

- Execution time
  - Asur ~5.5h
  - Palla ~7 days
  - **GED ~4h**

- Group extraction method
  - Palla works **only** with **CPM**
  - Asur and GED work with **any group extraction method**
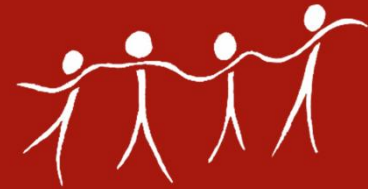
Experiments: Results

- Palla returns **all** possible events between groups, but does **not** assign its **type**

- Asur does **not** return **all** events and sometimes assigns **many events** (overlapping groups)

- GED may return **all** events depending on α and β (near to 0) and assigns the event type
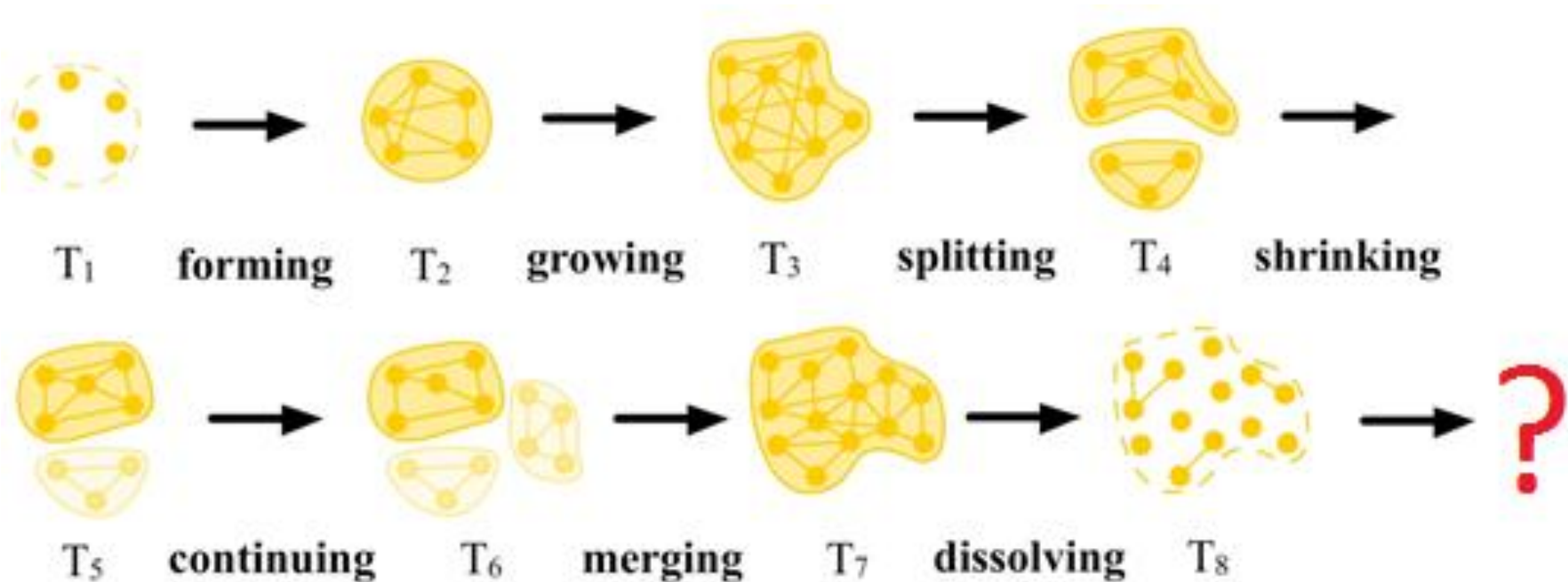
Final Remarks

- Identification of event types for group evolution

- Inclusion measure used for event discovery

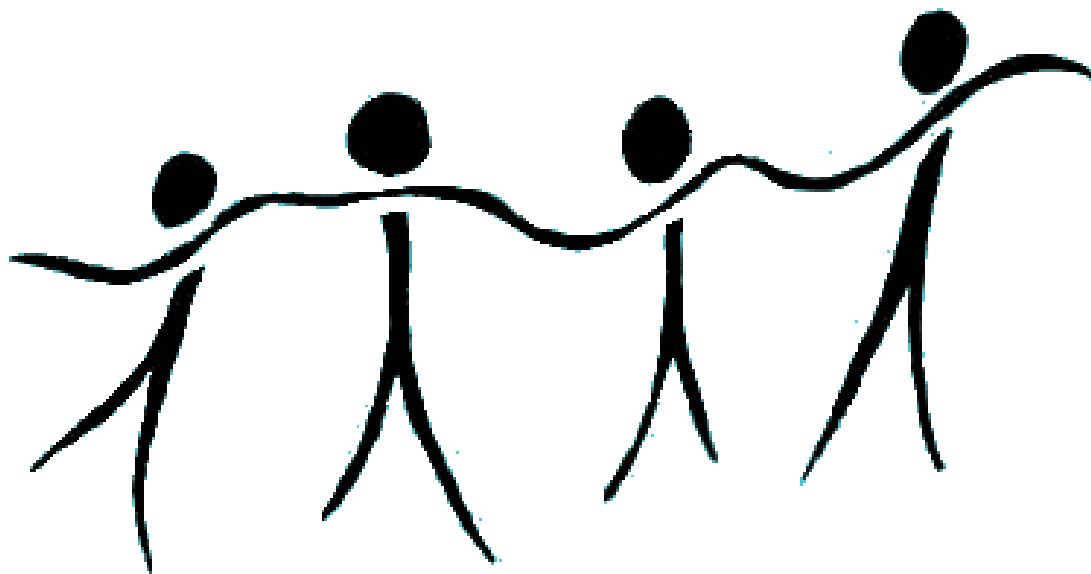- Group Evolution Discovery (GED)
  – a new method

Future Work: Event Prediction

T_1 **forming** T_2 **growing** T_3 **splitting** T_4 **shrinking**

T_5 **continuing** T_6 **merging** T_7 **dissolving** T_8

Thank you for your attention ☺

Social Network Group
@ Wroclaw University of Technology

Basic concepts: Social position

$$SP_{n+1}(x) = (1 - \varepsilon) + \varepsilon \cdot \sum_{y \in V} SP_n(y) \cdot C(y \to x)$$

Aktor	SP	Rank SP	CD	Rank CD
A	0,566	4	2	4
B	0,667	5	3	2
C	1,440	1	4	1
D	1,217	2	2	4
E	1,110	3	3	2