COMMUNITY DETECTION IN SIGNED SOCIAL NETWORKS USING VARIABLE LENGTH GENETIC ALGORITHM

PRESENTED BY:
Manisha Keim

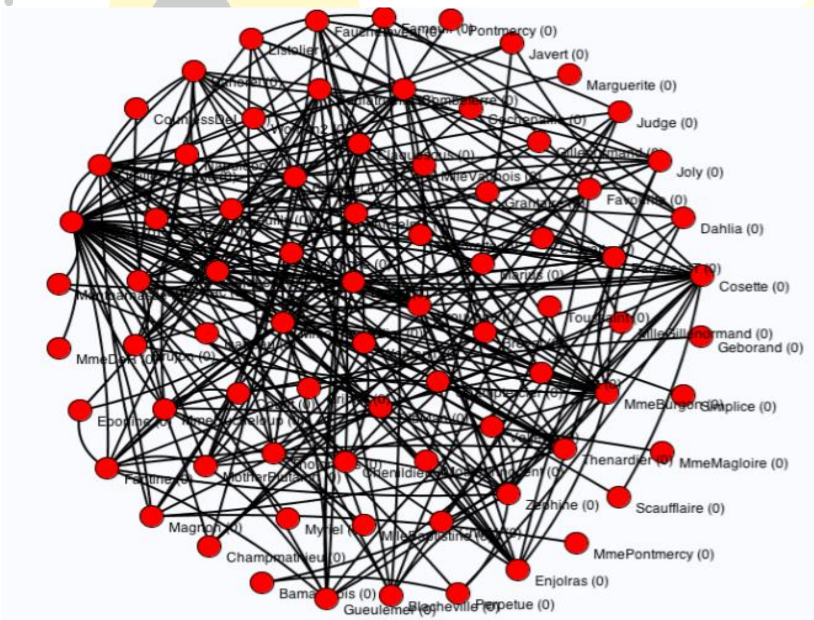
SUPERVISOR:
Dr. Vinti Agarwal

OUTLINE

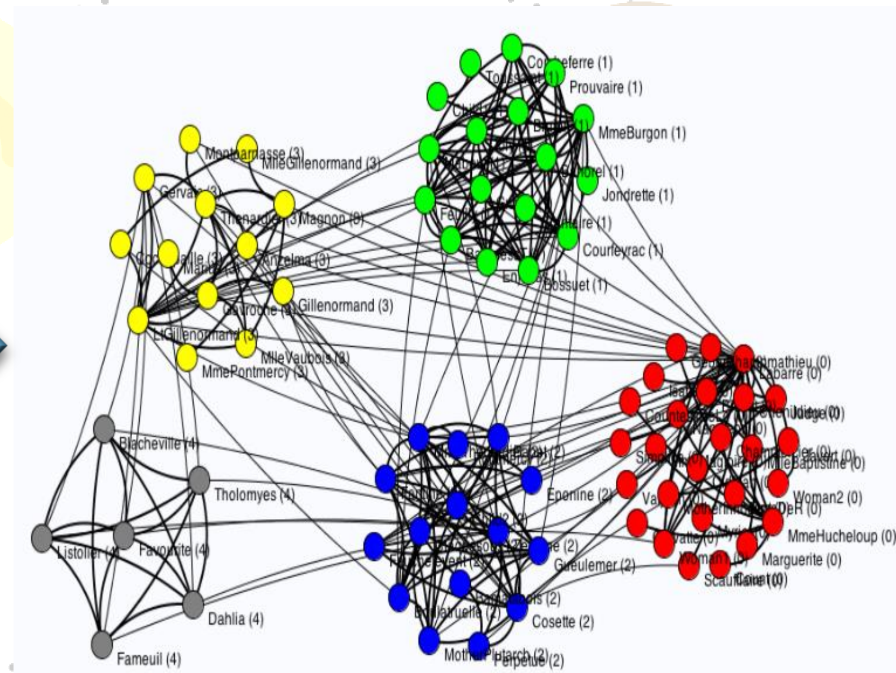
- Community Detection in Signed Networks
- Datasets : Extended Epinions, Slashdot
- Theories in Signed Networks
- Consensus Evaluation
- Variable Length Genetic Algorithm
- Proposed Work
- References

COMMUNITY DETECTION

- COMMUNITY DETECTION: The problem that community detection attempts to solve is the identification of groups of vertices that are **more densely connected** to each other than to the rest of the network.

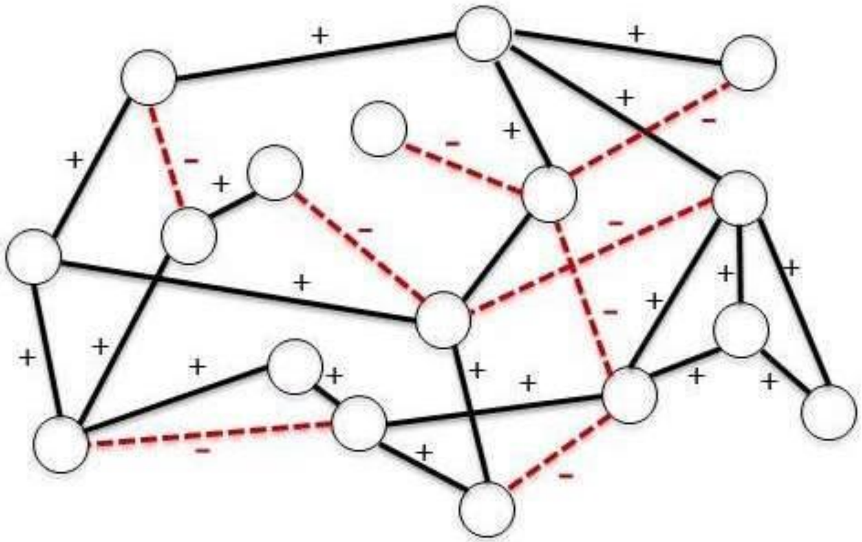


BEFORE



AFTER

SIGNED NETWORKS



Extended Epinions : Trust / Distrust

Slashdot : Friend / foe

DATASETS : EXTENDED EPINIONS, SLASHDOT

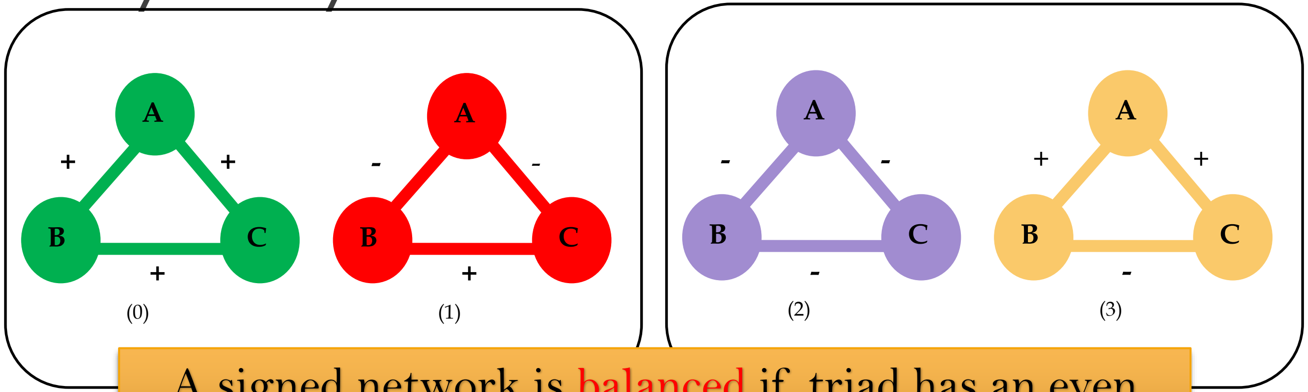
EXTENDED EPINION	SLASHDOT
Trust / Distrust	Friend / foe
~132,000 users and 841,372 statements	77,360 users and 905,468 edges
Directed	Directed
Users and Items are represented by anonimized numeric identifiers.	(u,v) : u's approval or disapproval of v's comments

THEORIES IN SIGNED NETWORK

- **Social Balance Theory**
- **Social Status Theory**

SOCIAL BALANCE THEORY

- Basic idea is that persons seek to avoid ***tension*** or ***dissonance*** in their relations.
- Works on
  - Triads (groups of three)
  - Assumes only positive (+) or negative (-) relations
  - Assumes *symmetry* of relations

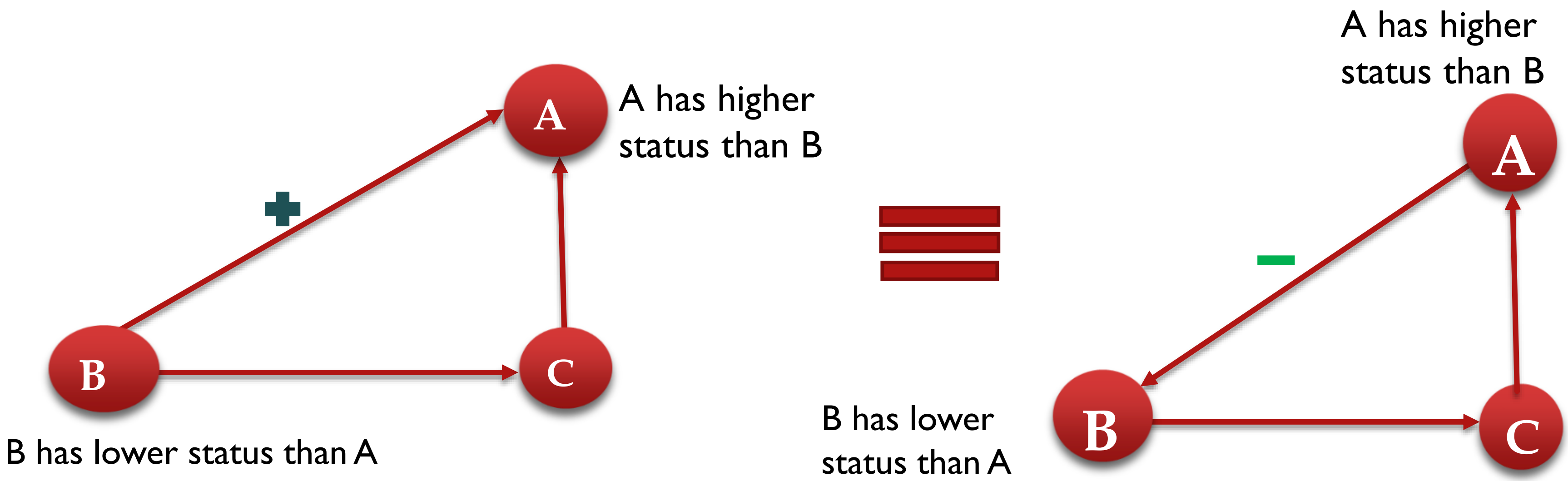


A signed network is balanced if triad has an even number of negative ties

ISIS brings Putin, Obama TOGETHER

SOCIAL STATUS THEORY

- Status theory [Leskovec et al.'10]
  - Link $u \xrightarrow{+} v$ means: v has higher status than u
  - Link $u \xrightarrow{-} v$ means: v has lower status than u

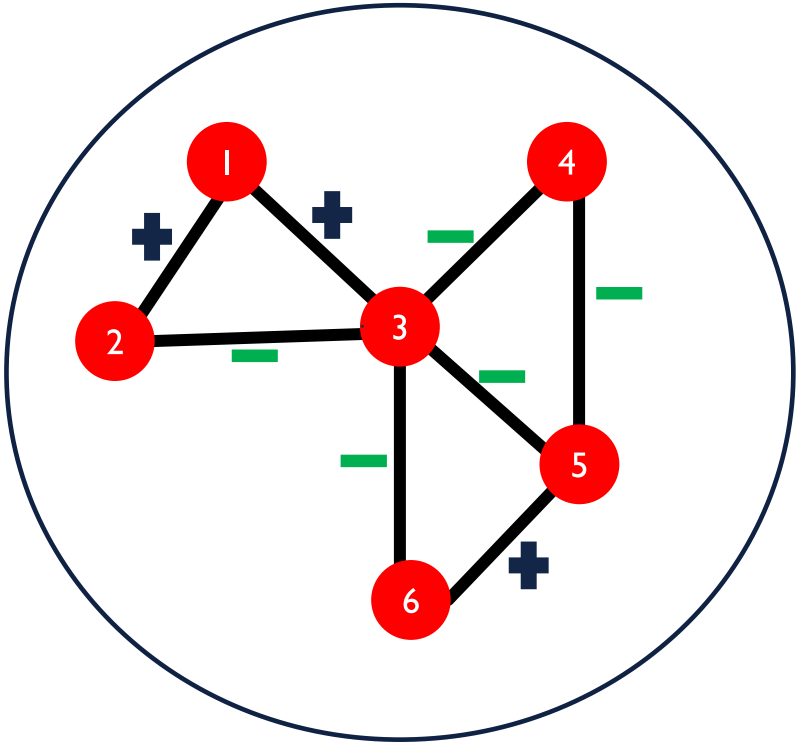


BALANCE INDEX(BI) COMPUTATION

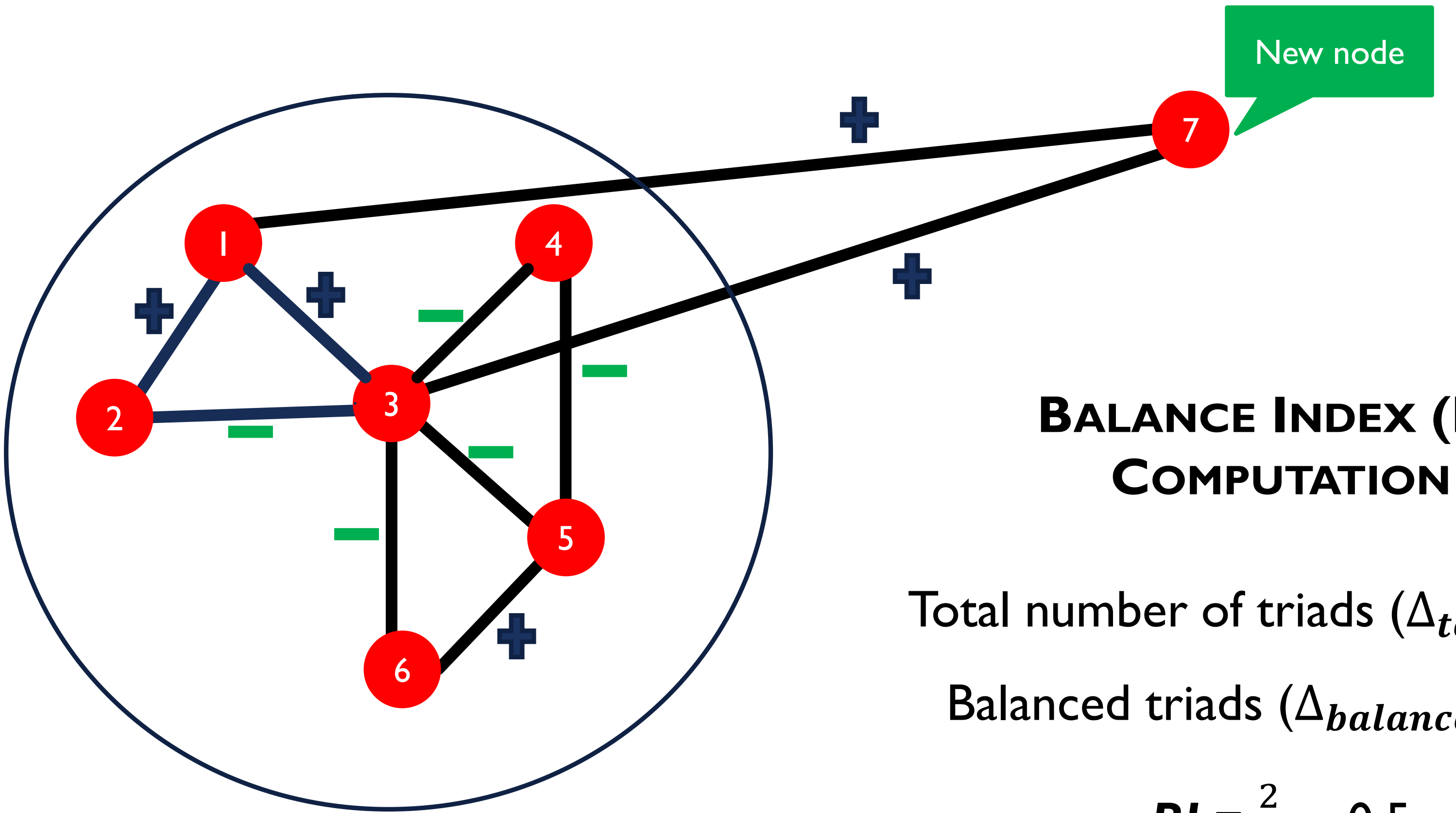
Total number of triads (Δ_{total}) = 3

Balanced triads ($\Delta_{balanced}$) = 1

$$\boldsymbol{BI} = \frac{1}{3} = 0.3333$$



BALANCE INDEX(BI) COMPUTATION



$BI_{new} > BI_{old}$

BALANCE INDEX (BI) COMPUTATION

Total number of triads (Δ_{total}) = 4

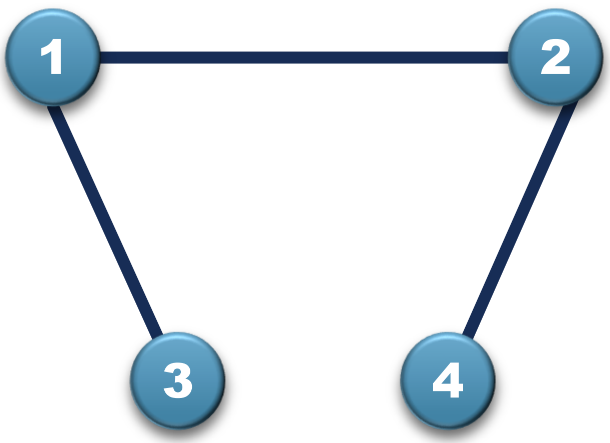
Balanced triads ($\Delta_{balanced}$) = 2

$\boldsymbol{BI} = \frac{2}{4} = 0.5$

Consensus Evaluation

- Fuzzy m-*ary* relations
  - Preference Relations
  - Fuzzy Binary Relations
- Trust/distrust relations
- Direction of edges.

BINARY ADJACENCY MATRIX



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

<u>LIMITATIONS OF BINARY RELATIONS :</u>

- ❑ It can be used only to represent pair wise relations.
- ❑ Degree of relationship can't be defined because it used only crisp values that is either 0 or1.

PREFERENCE RELATIONS

- A preference relation P on a set $I = \{i_1, .., i_k\}$ is characterized by a function $\mu_P: I \times I \rightarrow [0,1]$, such that $\mu_P(i_q, i_q) = 0.5\ \forall q$ and $\mu_P(i_q, i_r) + \mu_P(i_r, i_q) = 1\ \forall q, r$.

$$\mu_P(i_q, i_r) \begin{cases} 1, & if\ i_q\ is\ definately\ preferred\ over\ i_r \\ \alpha \in\ ]0.5, 1[& if\ i_q\ is\ preferred\ over\ i_r \\ 0.5, & if\ there\ is\ indifference\ between\ i_q\ and\ i_r \\ \beta \in\ ]0, 0.5\ [& if\ i_r\ is\ preferred\ over\ i_q \\ 0, & if\ i_r\ is\ definately\ preferred\ over\ i_q \end{cases}$$

PREFERENCE RELATIONS

Let us assume there are 3 items I_1, I_2, I_3 and users have given preferences on them.

ITEMS → I_1 I_2 I_3

USER PREFERENCES → 2 3 1

$$\mu(I_1, I_2) =]0, 0.5[$$

$$\mu(I_2, I_3) =]0.5, 1[$$

EXAMPLE : FUZZY PREFERENCE RELATIONS

Let us assume there are 6 users $d_1, d_2, d_3, d_4, d_5, d_6$ who has given their preferences on 3 items I_1, I_2, I_3.

$$\mathbf{P}^{(1)} = \begin{pmatrix} 0.5 & 0.9 & 1 \\ 0.1 & 0.5 & 0.6 \\ 0 & 0.4 & 0.5 \end{pmatrix} \quad \mathbf{P}^{(2)} = \begin{pmatrix} 0.5 & 0.8 & 1 \\ 0.2 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

$$\mathbf{P}^{(3)} = \begin{pmatrix} 0.5 & 0.9 & 0.9 \\ 0.1 & 0.5 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} \quad \mathbf{P}^{(4)} = \begin{pmatrix} 0.5 & 1 & 0.7 \\ 0 & 0.5 & 0.7 \\ 0.3 & 0.3 & 0.5 \end{pmatrix}$$

$$\mathbf{P}^{(5)} = \begin{pmatrix} 0.5 & 0.1 & 0.1 \\ 0.9 & 0.5 & 0 \\ 0.9 & 1 & 0.5 \end{pmatrix} \quad \mathbf{P}^{(6)} = \begin{pmatrix} 0.5 & 0.5 & 0.6 \\ 0.5 & 0.5 & 0.9 \\ 0.4 & 0.1 & 0.5 \end{pmatrix}$$

FUZZY BINARY ADJACENCY RELATION

- A fuzzy binary relation R_2 representing the degree of relationship exist between d_i and d_j, is defined using the member function $\mu_{R_2}: D \times D \to [0, 1]$ as:

$$\mu_{R_2}(d_i, d_j) = \begin{cases} 1, & if\, d_i \; is\; strongly\; connected\; with\; d_j \\ \gamma \in \,]0,1[& d_i\; is, to\; some\; extent, related\; with\; d_j \\ 0, & if\, d_i\; is\; not\; related\; to\; d_j \end{cases}$$

FORMULA

$$\mu_{R_2}(d_i, d_j) = 1 - \frac{2}{k(k-1)} \sum_{\substack{q,r=1 \\ q<r}}^{k} \left| P_{qr}^{(i)} - P_{qr}^{(j)} \right|$$

EXAMPLE : FUZZY BINARY ADJACENCY RELATIONS

The next step consists in deriving the degrees of pairwise adjacency,

$$\mu_{R_2}(d_1, d_2) = 1 - \frac{1}{3}\sum_{\substack{q,r=1\\q<r}}^{3}\left|p_{qr}^{(1)} - p_{qr}^{(2)}\right| \approx 0.93$$

$$\mu_{R_2}(d_1, d_3) = 1 - \frac{1}{3}\sum_{\substack{q,r=1\\q<r}}^{3}\left|p_{qr}^{(1)} - p_{qr}^{(3)}\right| \approx 0.96$$

$$\cdots = \cdots$$

$$\mu_{R_2}(d_4, d_6) = 1 - \frac{1}{3}\sum_{\substack{q,r=1\\q<r}}^{3}\left|p_{qr}^{(4)} - p_{qr}^{(6)}\right| \approx 0.73$$

$$\mu_{R_2}(d_5, d_6) = 1 - \frac{1}{3}\sum_{\substack{q,r=1\\q<r}}^{3}\left|p_{qr}^{(5)} - p_{qr}^{(5)}\right| \approx 0.4$$

EXAMPLE : FUZZY BINARY ADJACENCY RELATIONS

Such values of pairwise adjacency can be conveniently collected into the following matrix representing the fuzzy binary adjacency relation.

$$\mathbf{R} \approx \begin{pmatrix} 1 & 0.933 & 0.967 & 0.833 & 0.233 & 0.633 \\ 0.933 & 1 & 0.9 & 0.767 & 0.3 & 0.633 \\ 0.967 & 0.9 & 1 & 0.866 & 0.267 & 0.667 \\ 0.833 & 0.767 & 0.866 & 1 & 0.267 & 0.73 \\ 0.233 & 0.3 & 0.267 & 0.267 & 1 & 0.4 \\ 0.633 & 0.633 & 0.667 & 0.73 & 0.4 & 1 \end{pmatrix}.$$

FUZZY M-ARY RELATIONS

- A fuzzy m-ary relation R_m on the set D is defined by the membership function $\mu_{R_m}: D^m \rightarrow [0, 1]$,

$$\mu_{R_m}(d_{p_1}, \ldots, d_{p_m}) = \begin{cases} 1, & if\, d_{p_1}, \ldots, d_{p_m} \text{ are strongly related with each other} \\ \gamma \in]0, 1[& if\, d_{p_1}, \ldots, d_{p_m} \text{ are, to some extent, related with each other} \\ 0, & if\, d_{p_1}, \ldots, d_{p_m} \text{ are definately not related with each other} \end{cases}$$

FORMULA

$$\mu_{R_m}(d_{p_1}, \ldots, d_{p_m}) = \frac{1}{\binom{m}{2}} (\mu_{R_2}(d_{p_1}, d_{p_2}) + \cdots \ldots \ldots . . + \mu_{R_2}(d_{p_{m-1}}, d_{p_m}))$$

FUZZY M-ARY RELATIONS

$$\mu_{R_3}(d_1, d_2, d_3) = \frac{1}{3}\mu_{R_2}(d_1, d_2) + \frac{1}{3}\mu_{R_2}(d_1, d_3) + \frac{1}{3}\mu_{R_2}(d_2, d_3) \approx 0.93$$

$$\mu_{R_3}(d_3, d_5, d_6) = \frac{1}{3}\mu_{R_2}(d_3, d_5) + \frac{1}{3}\mu_{R_2}(d_3, d_6) + \frac{1}{3}\mu_{R_2}(d_5, d_6) \approx 0.4$$

Since $\mu_{R_3(d_3,d_5,d_6)} < \mu_{R_3(d_1,d_2,d_3)}$
Therefore, members of the group (d_1, d_2, d_3) are highly similar than the members of group (d_3, d_5, d_6)

GENETIC ALGORITHM - DEFINITION

"A method for moving from one population of "chromosomes" to a new population by using a kind of "natural selection" together with the genetics‒inspired operators of crossover, mutation, and inversion".

- John Holland, 1975

"Genetic Algorithms are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and natural genetics".

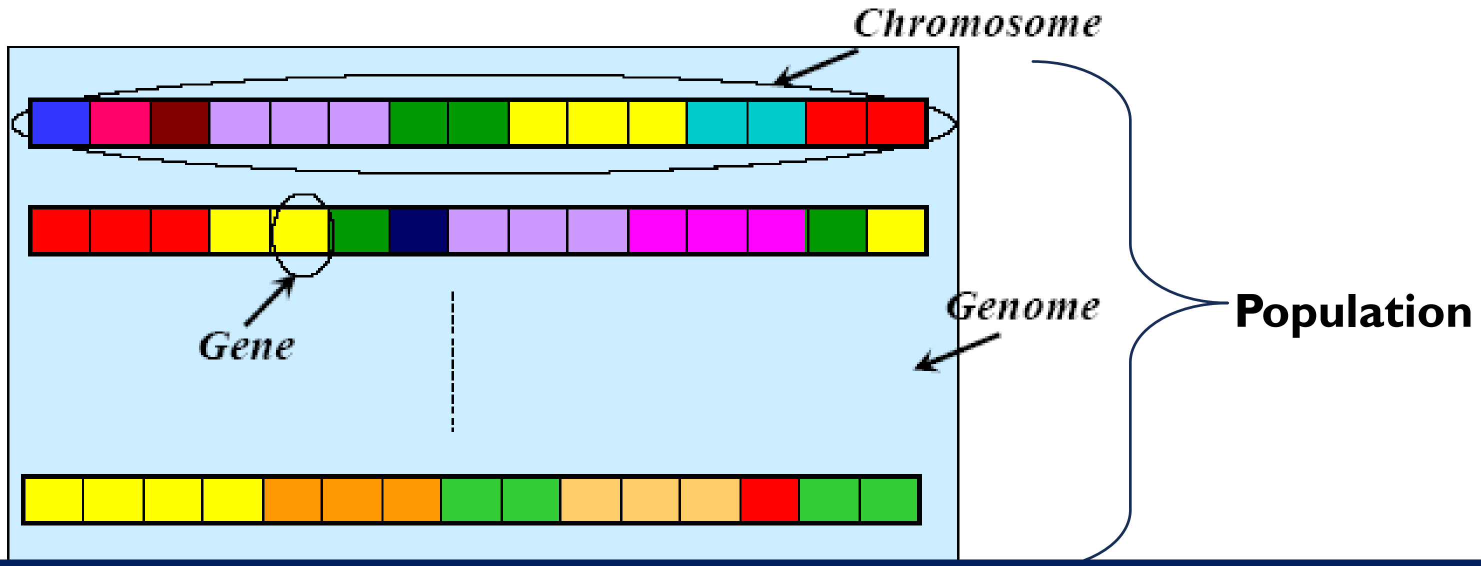
\- David E. Goldberg, 1989

GENETIC ALGORITHM INCLUDES

- Chromosome
- Operators
- Objective Function
- Stopping Criteria

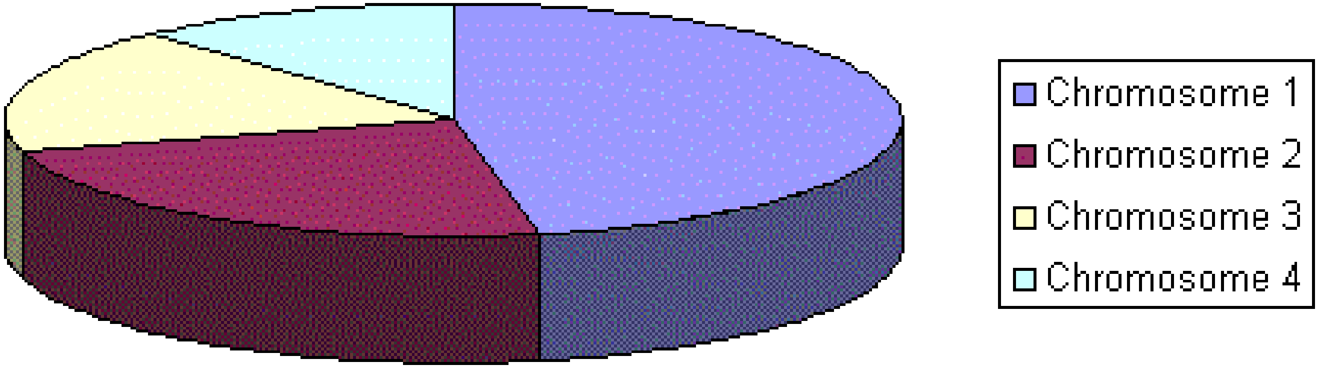
KEY TERMS

- **Chromosome** : a chromosome (also sometimes called a genotype) is a set of parameters which define a proposed solution to the problem.



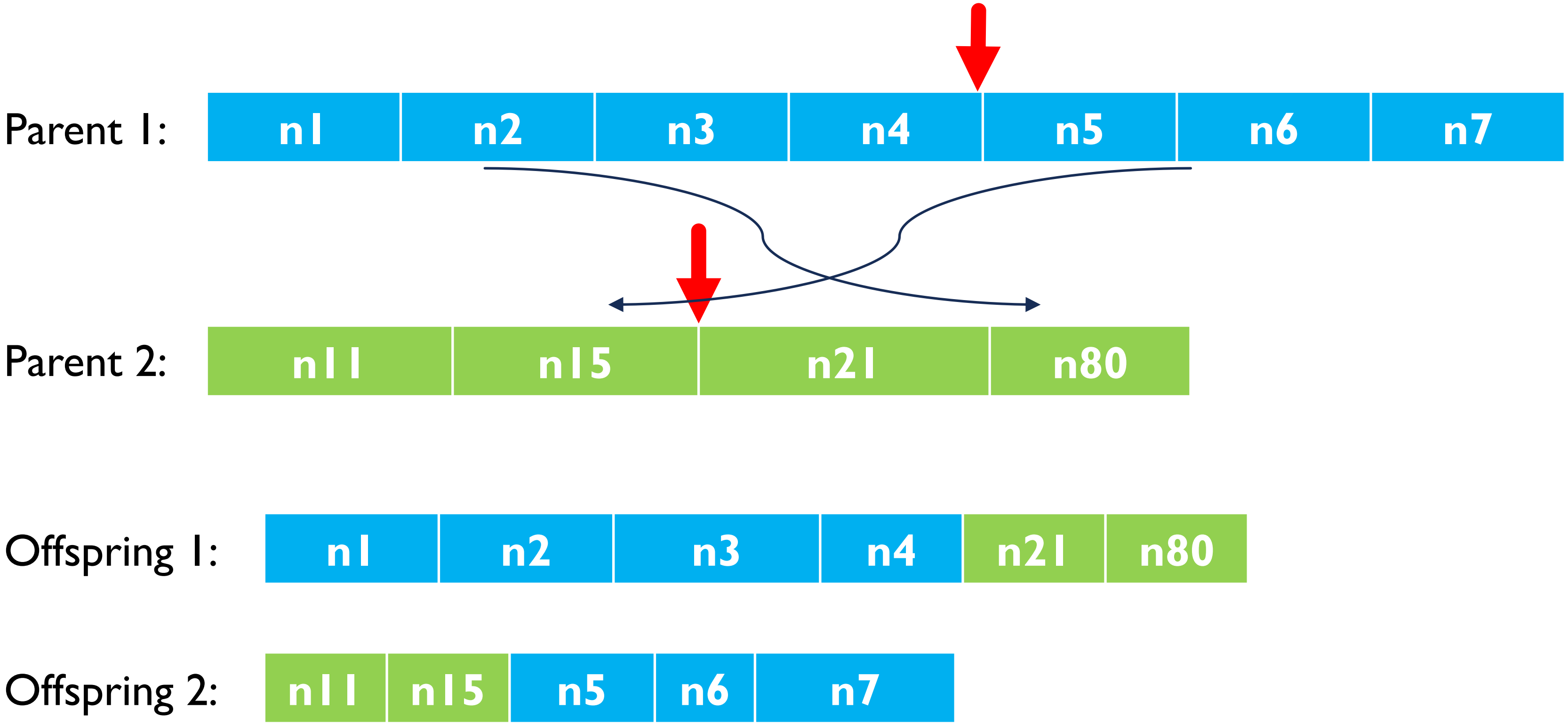
SELECTION

- <u>Selection</u>: replicates the most successful solution found in a population.
- **Roulette−wheel Sampling**, which is conceptually equivalent to giving each individual a slice of a circular roulette wheel equal in area to the individual's fitness. The roulette wheel is spun, the ball comes to rest on one wedge−shaped slice, and the corresponding individual is selected.

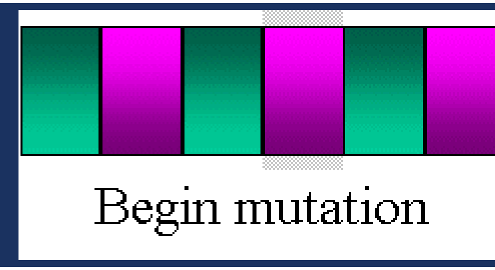


CROSSOVER

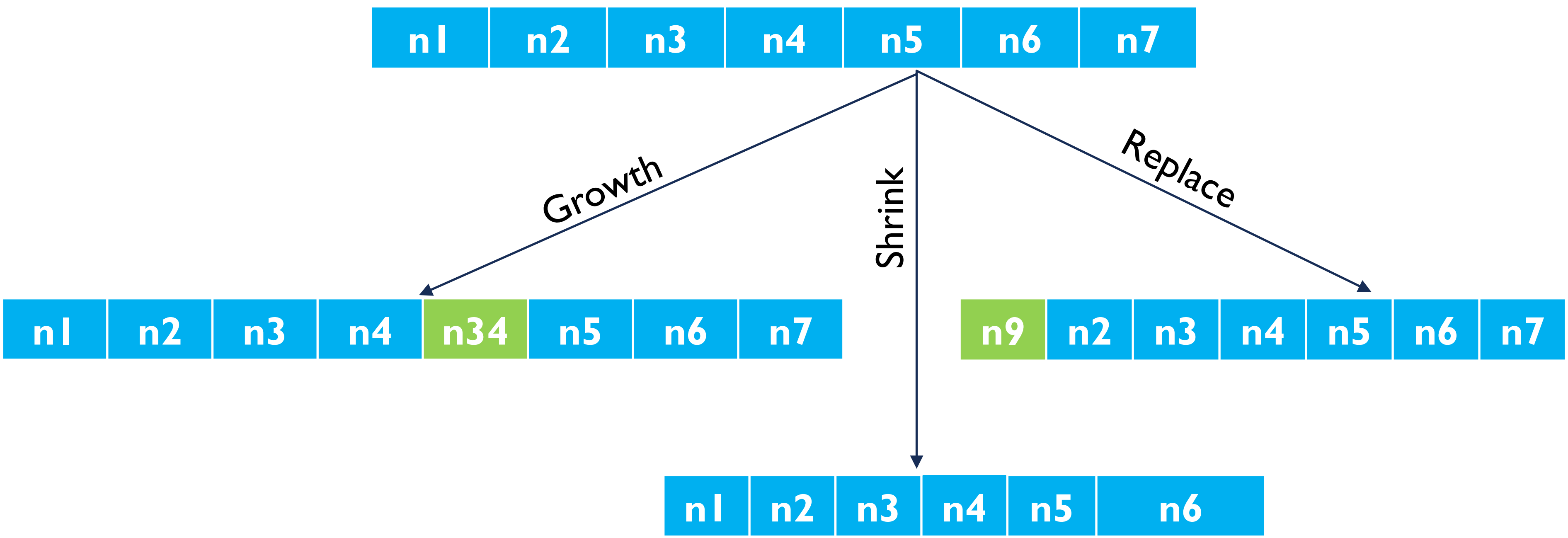
Crossover : decomposes two distinct solutions and then randomly mixes their parts to form new solutions



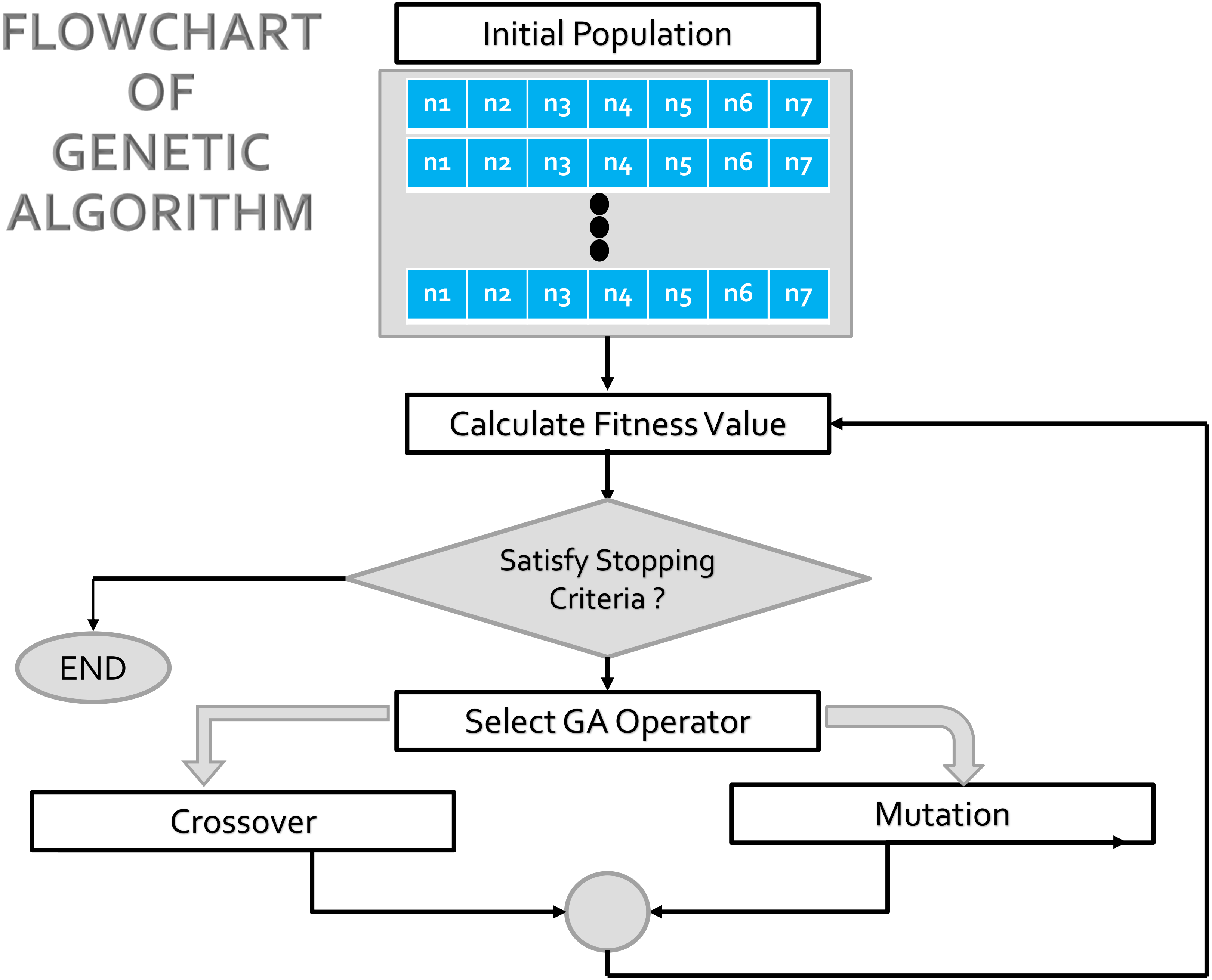
MUTATION



Mutation: randomly changes a candidate solution.



FLOWCHART OF GENETIC ALGORITHM



FITNESS FUNCTION

- Fitness function validates the diversion of process towards its optimization goal by allowing best individuals to breed that lead to good recommendation

- Objective function = $\frac{2 \times Consensus \times Balance\ Index}{Balance\ Index + Consensus}$

- We want high value of Consensus and Balance Index to get a high value of objective function.

STOPPING CRITERIA

- When a maximum number of generation elapsed or
- a desired level of fitness value is achieved.

PROPOSED WORK

- Community Detection in Signed Networks by modeling both positive and negative relationships among users using variable length genetic algorithm.
- By computing various social factors (i.e. balance index, social status) which helps in avoiding the possible imbalance arise in communities.
- Incorporate consensus building among a group of networked decision makers using *Fuzzy m-ary* relationships (Brunelli, *et al*., 2014).

REFERENCES

- J.Leskovec, D.Huttenlocher, J.Kleinberg, **Predicting Positive and Negative Links in Online Social Networks**,ACM CHI(2010 April).
- P.Esmailian and M.Jalili, **Community Detection in Signed Networks : The role of negative ties in different scales.**
- M.Brunelli, M.Fedrizzi and Mi.Fedrizzi, **Fuzzy m-ary adjacency relations in social network analysis : Optimization and consensus evaluation.**
- Patidar, A., Agarwal, V., Bharadwaj, K.K. , **Predicting Friends and Foes in Signed Networks using Inductive Inference and Social Balance Theory.**
- Agarwal, V., Bharadwaj, K.K. (2015) **Recommending Diverse Friends in Signed Social Networks based on Adaptive Soft Consensus Paradigm Utilizing Fuzzy m-ary Adjacency Relations.**
- Lee, C.Y., Antonsson, E.K.: **Variable Length Genomes for Evolutionary Algorithms.** In L. Darrell Whitley; David E. Goldberg; Erick Cantú-Paz; Lee Spector; Ian C. Parmee& Hans-Georg Beyer, ed., 'GECCO' , Morgan Kaufmann, , pp. 806 (2009)
- Kim, I.Y., de Weck, O.L.: **Variable chromosome length genetic algorithm for progressive refinement in topology optimization**. Struct. Multidisc. Optim. 29, 445–456 (2005).
- **An Introduction to Genetic Algorithms**, Melanie Mitchell, MIT Press, 1999

THANK YOU