BIG DATA ANALYTICS

A Social Network Approach

Telkom University

Andry Alamsyah

Social Computing and Big Data Research Group

Fakultas Ekonomi dan Bisnis

Andry Alamsyah

Fakultas Ekonomi dan Bisnis

Universitas Telkom

Research Field :

Social Network, Complex Network / Network Science, Social Computing, Data Analytics, Data Mining, Big Data, Graph Theory, Content Business, Data Business, ICT Business

andry.alamsyah@gmail.com

andrya.staff.telkomuniversity.ac.id

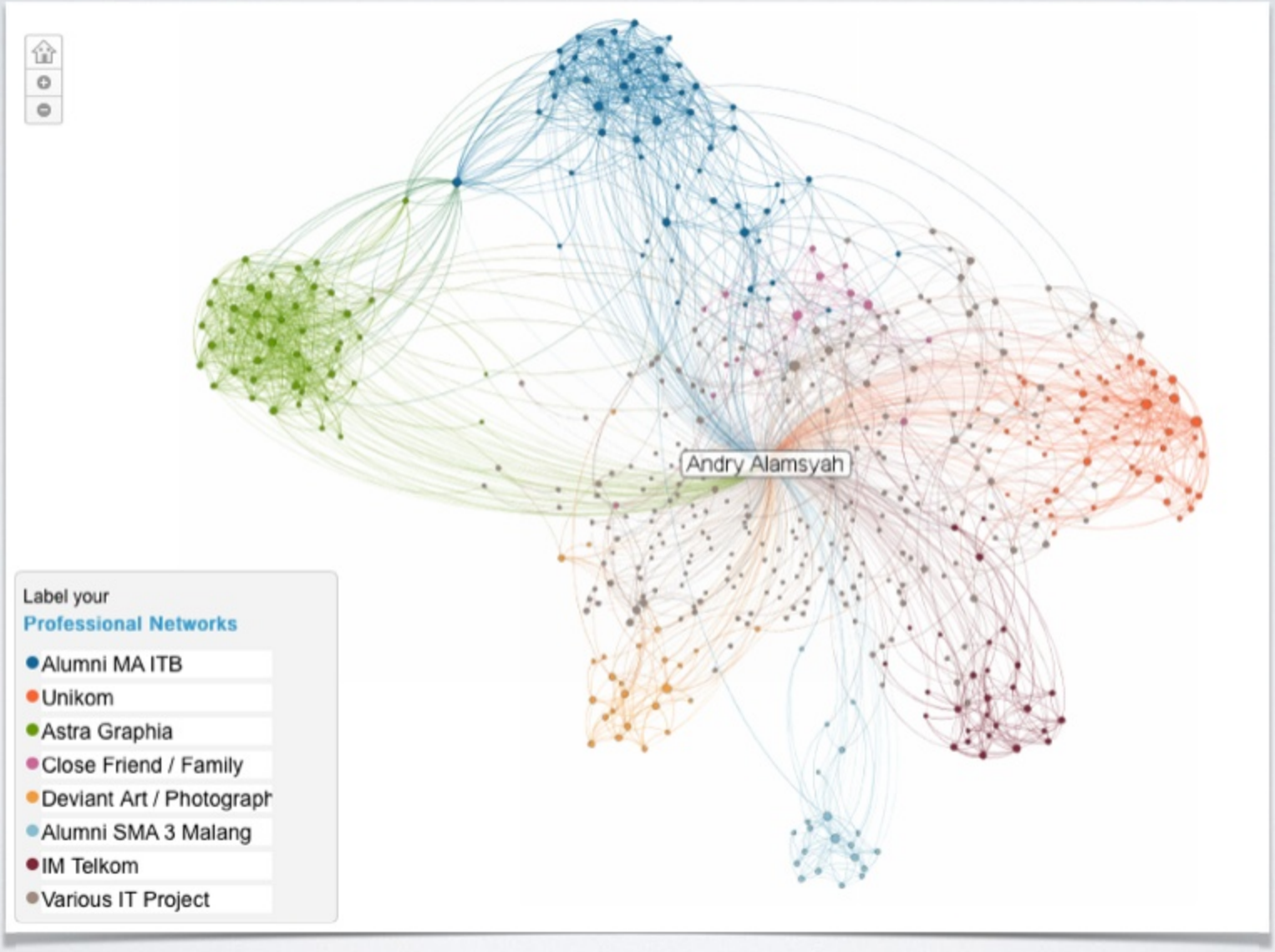
telkomuniversity.academia.edu/andryalamsyah

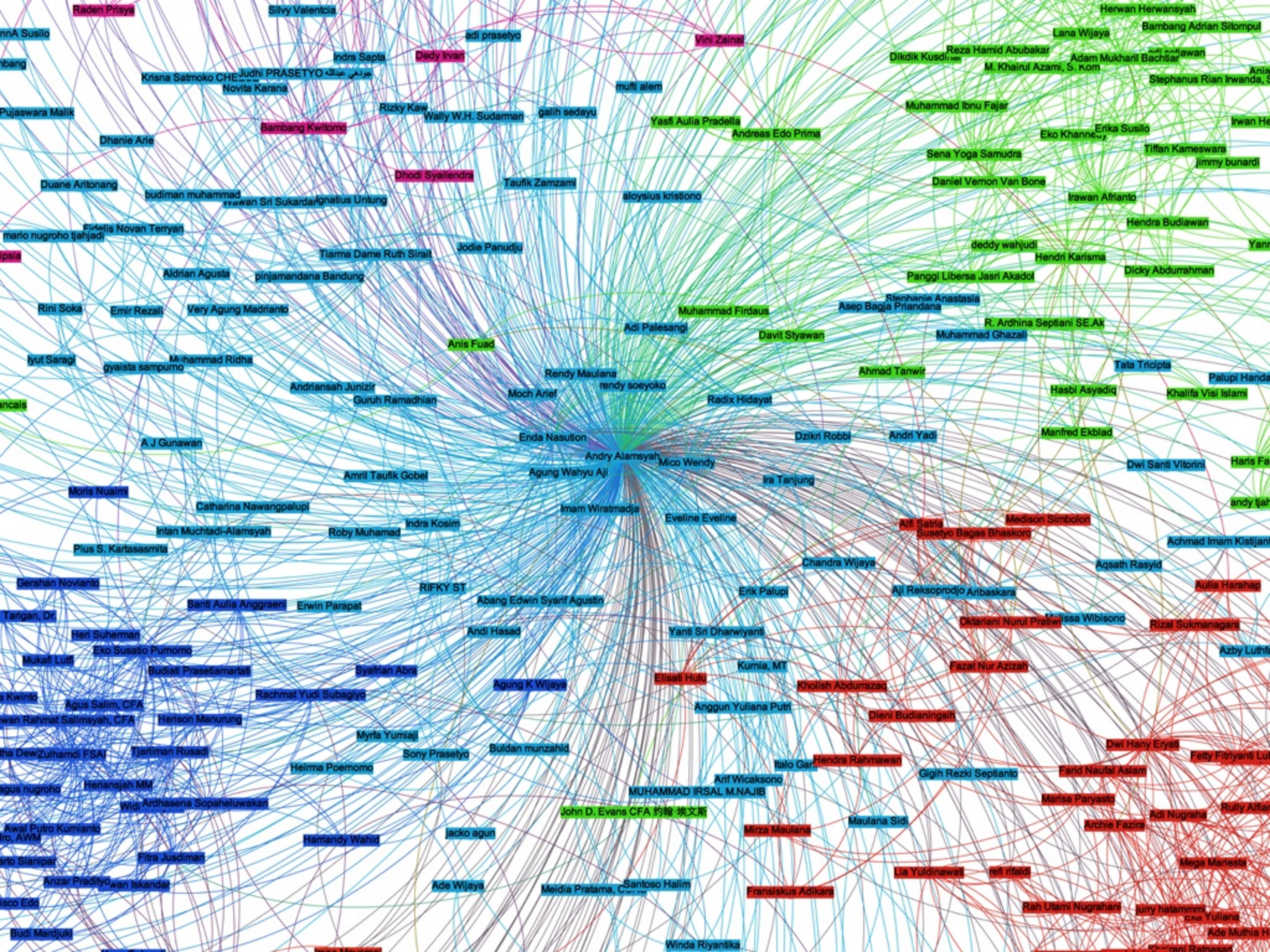
researchgate.net/profile/Andry_Alamsyah

linkedin.com/andry.alamsyah

twitter.com/andrybrew

WHO AM I ?

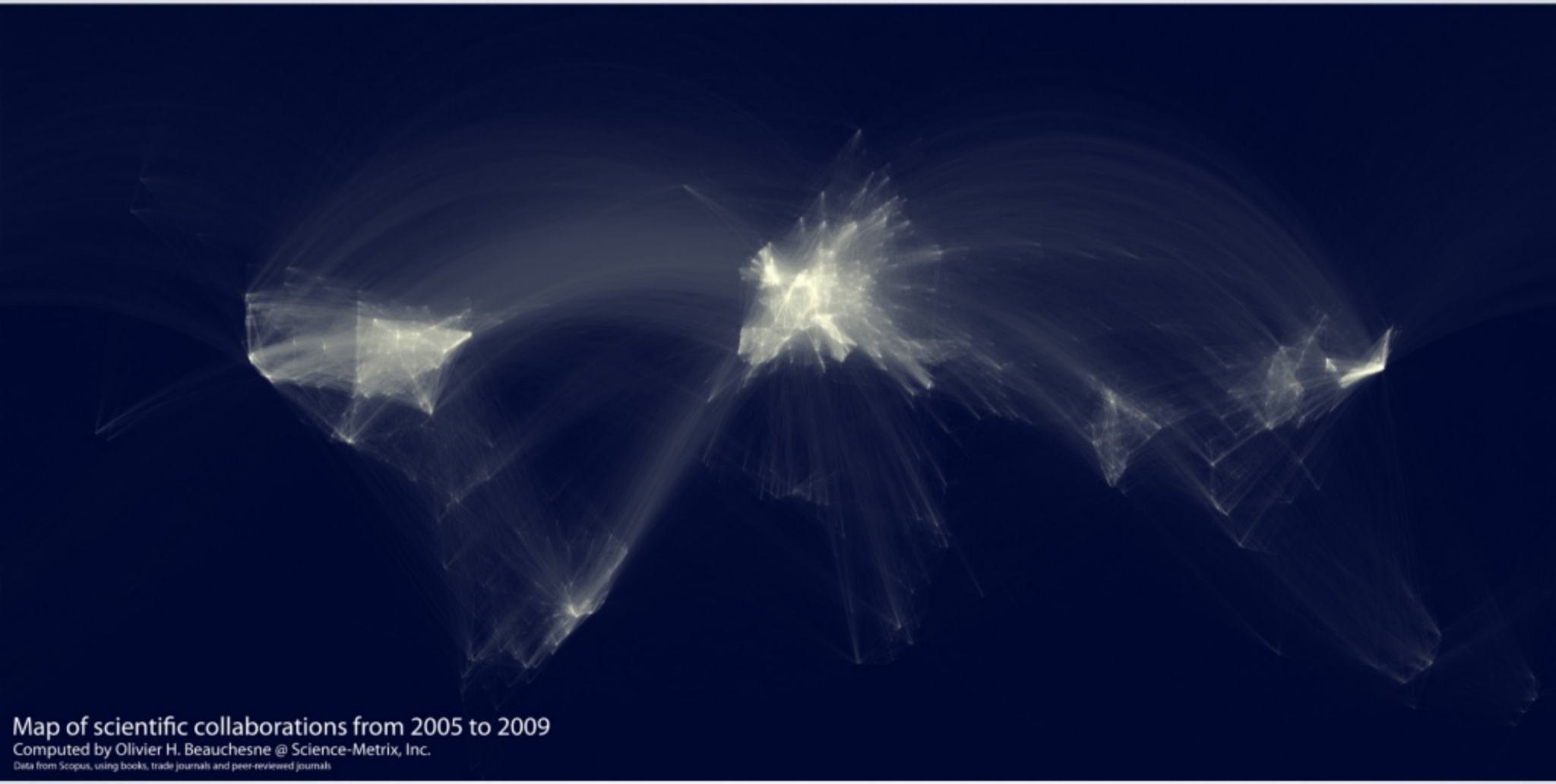

Andry Alamsyah
Label your
Professional Networks
Alumni MA ITB
Unikom
Astra Graphia
Close Friend / Family
Deviant Art / Photograph
Alumni SMA 3 Malang
IM Telkom
Various IT Project


Raden Prisya
Silvy Valentcia
Herwan Herwansyah
Bambang Adrian Sitompul
Lana Wijaya
adi prasetyo
Vini Zainal
Indra Sapta
Dedy Irvan
Reza Hamid Abubakar
Adam Mukharil Bachtiar
M. Khairul Azami, S. Kom
Judhi PRASETYO
Novita Karana
mufti alem
Muhammad Ibnu Fajar
Rizky Kaw
Wally W.H. Sudarman
galih sedayu
Yasfi Aulia Pradella
Bambang Kwitomo
Andreas Edo Prima
Erika Susilo
Dhanie Arie
Sena Yoga Samudra
Tiffan Kameswara
jimmy bunardi
Dhodi Syailendra
Taufik Zamzami
Duane Aritonang
Daniel Vernon Van Bone
budiman muhammad
Ignatius Untung
aloysius kristiono
Irawan Afrianto
Hendra Budiawan
mario nugroho tjahjadi
Jodie Panudju
deddy wahjudi
Tiarma Dame Ruth Sirait
Hendri Karisma
Aldrian Agusta
pinjamandana Bandung
Dicky Abdurrahman
Panggi Libersa Jasri Akadol
Rini Soka
Emir Rezali
Very Agung Madrianto
Muhammad Firdaus
Asep Bagja Priandana
R. Ardhina Septiani SE.Ak
Adi Palesangi
Davit Styawan
Muhammad Ghazali
Anis Fuad
Tata Tricipta
Ahmad Tanwir
Rendy Maulana
rendy soeyoko
Andriansah Junizir
Hasbi Asyadiq
Khalifa Visi Islami
Moch Arief
Guruh Ramadhian
Radix Hidayat
Enda Nasution
Dzikri Robbi
Andri Yadi
Manfred Ekblad
A J Gunawan
Andry Alamsyah
Mico Wendy
Dwi Santi Vitorini
Amril Taufik Gobel
Agung Wahyu Aji
Ira Tanjung
Moris Nuaimi
Catharina Nawangpalupi
Imam Wiratmadja
Eveline Eveline
Indra Kosim
Alfi Satria
Medison Simbolon
Intan Muchtadi-Alamsyah
Roby Muhamad
Susetyo Bagas Bhaskoro
Pius S. Kartasasmita
Chandra Wijaya
Aqsath Rasyid
Gerahan Novianto
RIFKY ST
Erik Palupi
Aji Reksoprodjo Aribaskara
Santi Aulia Anggraeni
Abang Edwin Syarif Agustin
Erwin Parapat
Melissa Wibisono
Oktariani Nurul Pratiwi
Heri Suherman
Andi Hasad
Yanti Sri Dharwiyanti
Eko Susatio Purnomo
Mukafi Lutfi
Budiati Prasetiamartati
Syafrian Abra
Kumia, MT
Elisati Hulu
Fazat Nur Azizah
Agung K Wijaya
Kholish Abdurrazaq
Rachmat Yudi Subagiyo
Anggun Yuliana Putri
Herison Manurung
Dieni Budianingsih
Myrta Yumiaj
Dwi Hany Eryati
Tjarliman Rusadi
Sony Prasetyo
Buldan munzahid
Hendra Rahmawan
Heirma Poernomo
Gigih Rezki Septianto
Farid Naufal Aslam
Arif Wicaksono
MUHAMMAD IRSAL M.NAJIB
Marisa Paryasto
Ardhasena Sopaheluwakan
John D. Evans CFA 約翰·埃文斯
Maulana Sidi
Adi Nugraha
Archie Fazira
Hamandy Wahid
Mirza Maulana
Fitra Jusdiman
Mega Marlesta
Anzar Praditya
Lia Yuldinawati
Ade Wijaya
Santoso Halim
Fransiskus Adikara
Rah Utami Nugrahani
Budi Mardjuki
Winda Riyantika

LARGE SCALE DATA

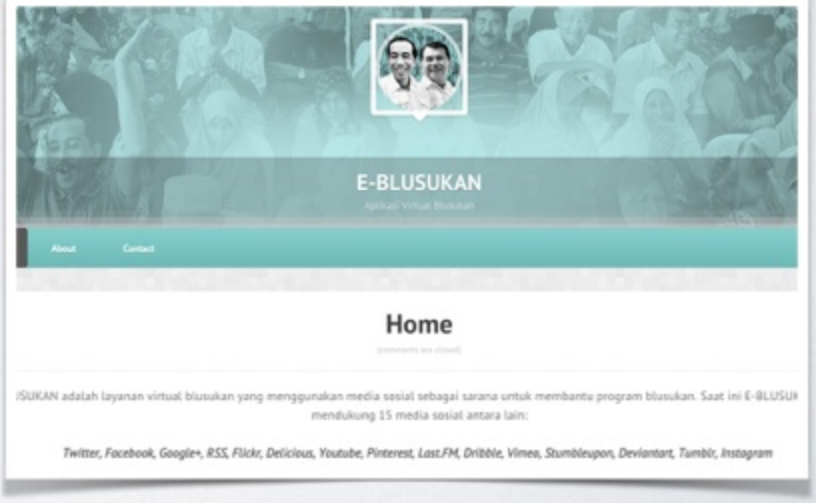
LARGE SCALE DATA

LARGE SCALE DATA



STORY / PHENOMENON

- BIG DATA leads to Social Computing *(Quantification of Individual / Social Behaviour)*
- Social Network Data / Conversation are widely available
- Social Network voices represent public voice become ***'Big'*** concern (references)
- The Need of Real-Time Analytic *(OLAP)*
- The Need of Powerful Metric for Social Network / Big Data



STORY / PHENOMENON

- There are many aspect of Big Data research, but too little resource, too little talent
- Same business objective, but increase effectiveness on top of current services
- Problem with Legacy Methodology approach using Questionnaire/ Interviews/Surveys *(ok with small scale data , expensive and took longer time for large scale data, accuracy issues)*

INDUSTRY EFFORTS

METHODS COMPARISON IN SOCIAL SCIENCE

LEGACY	DATA ANALYTICS
Confirmative	Explorative (Predictive)
Small Data Set	Larga Data Set
Small Number of Variable	Large Number of Variable
Deductive (no predictions)	Inductive
Numeric Data	Numeric and Non-Numeric Data
Clean Data	Data Cleaning

source : Data Mining and Statistics : Whats the Connections ? (Jerome Friedman)

BIG DATA STATE OF THE ART

Computation Related

Processing / Computation	Storage	Analytics Tools
• Hadoop • Nvidia CUDA • Twitter Storm • Bulk Synchronous Parallel Processing • GraphLab • Disk-Based Graph Processing	• neo4J • Titan • HDFS	• MLPACK • Mahout

Methodology / Analytics Related

modelling, descriptions, predictions, optimisation and simulation

Big Data - State of the art - Chalmers et al (2013)

BIG DATA ANALYTICS CONSTRUCTOR

Social Network

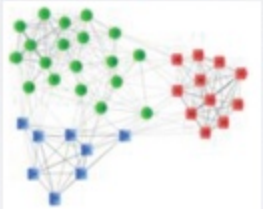
networks

tie-strength

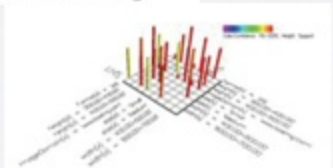
key players

cohesion

Data Mining



Clustering



Association Rules

Classification & Regression

Anomaly Detection

Sentiment Analysis

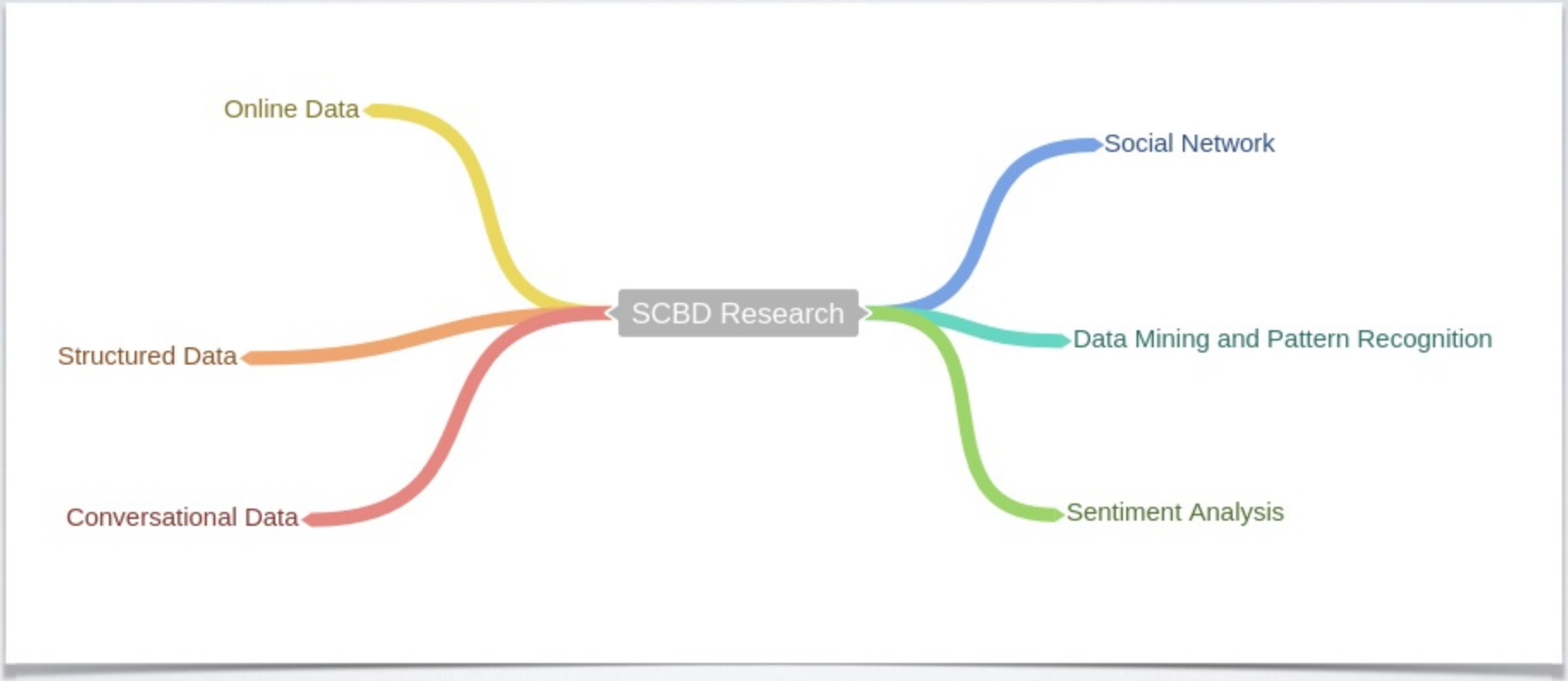
keyword spotting

lexical affinity

statistical methods

concept-level technique

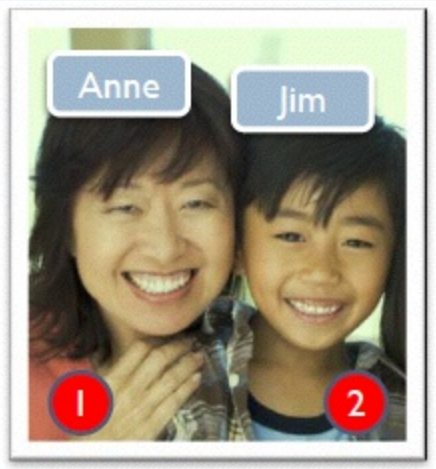
RESEARCH ROADMAP



GOAL : descriptions, predictions, optimisation and simulation

area : marketing, communications, knowledge management, operations, finance, etc

SOCIAL NETWORK MODEL



Can we study their interactions as a network ?

Communication

Anne: Jim, tell the Murrays they're invited

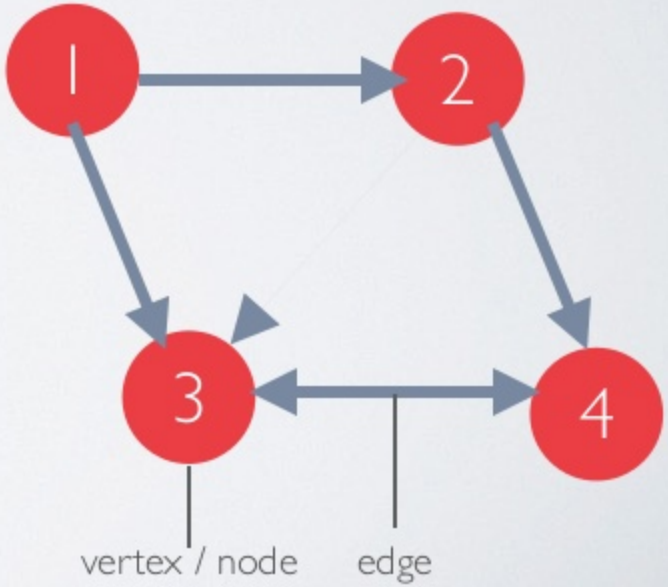
Jim: Mary, you and your dad should come for dinner!

Jim: Mr. Murray, you should both come for dinner

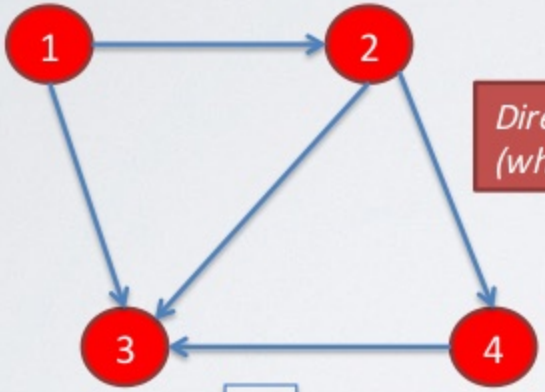
Anne: Mary, did Jim tell you about the dinner? You must come.

Mary: Dad, we are invited for dinner tonight

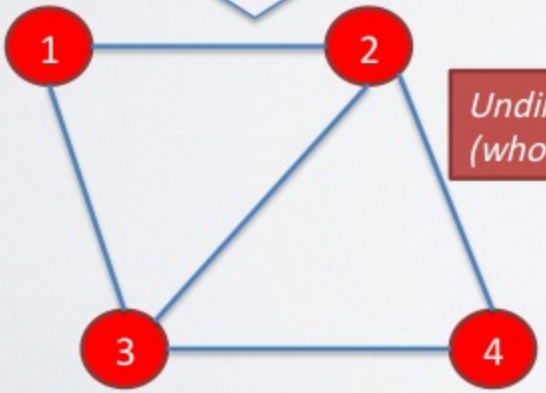
John: *(to Anne)* Ok, we're going, it's settled!



SOCIAL NETWORK MODEL



Directed
(who contact whom)



Undirected
(who knows whom)

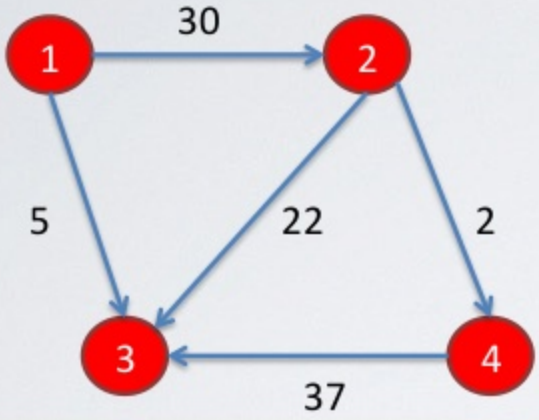
Edges List

Vertex	Vertex
1	2
1	3
2	3
2	4
4	3

Adjacency Matrix become symmetric

Vertex	1	2	3	4
1	-	1	1	0
2	1	-	1	1
3	1	1	-	0
4	0	1	0	-

TIE STRENGTH



Weight could be
- Frequency of interactions in period of observation
- Number of items exchanged in period
- Individual perceptions of strength of relationship
- Cost of communications or exchange, e.g. distance

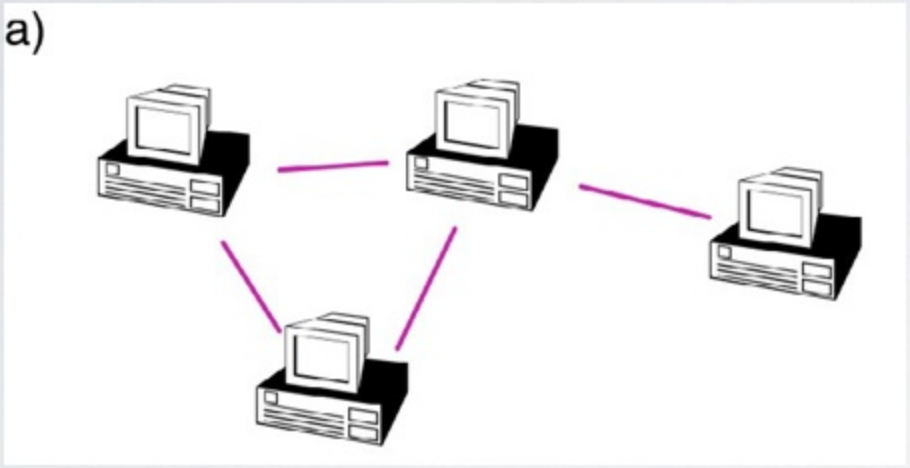
Edges List

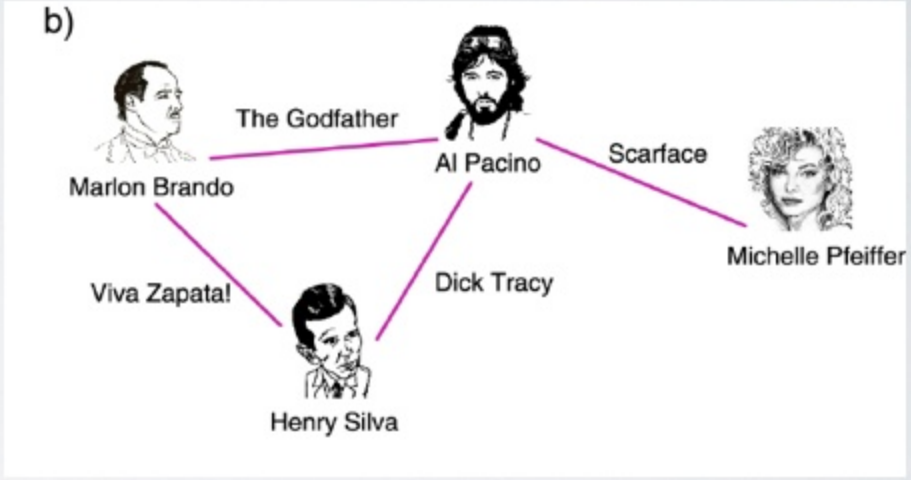
Vertex	Vertex	Weight
1	2	30
1	3	5
2	3	22
2	4	2
4	3	27

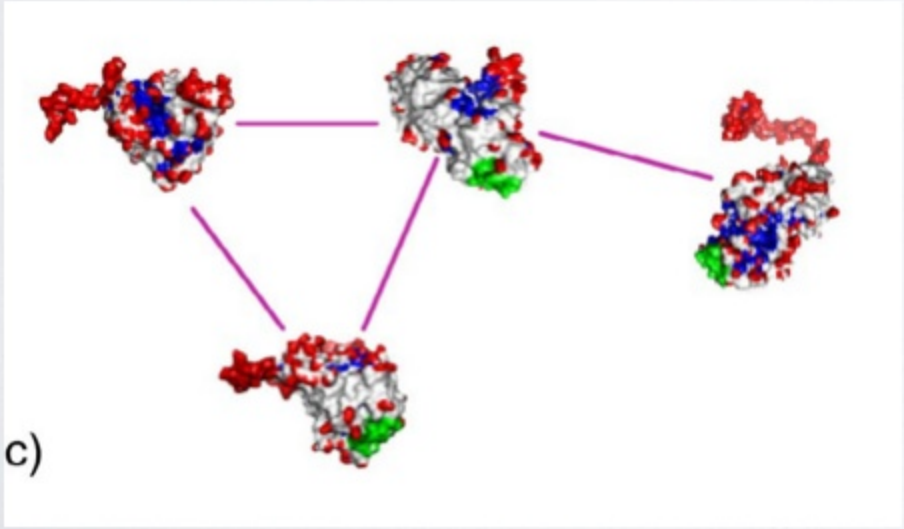
Adjacency Matrix (weight)

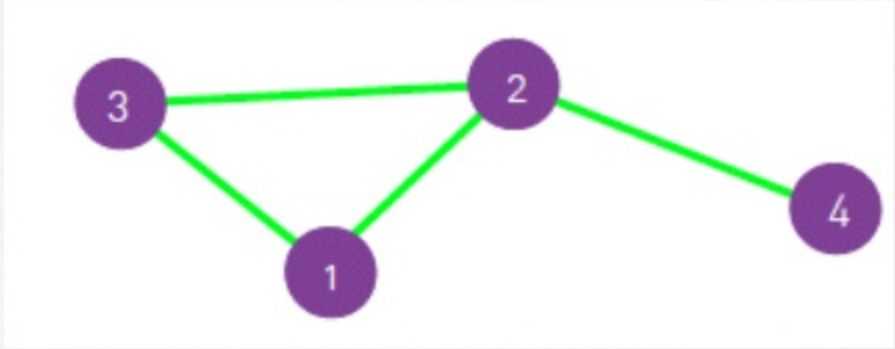
Vertex	1	2	3	4
1	-	30	5	0
2	30	-	22	2
3	5	22	-	37
4	0	2	37	-

NETWORK MODEL EXAMPLE



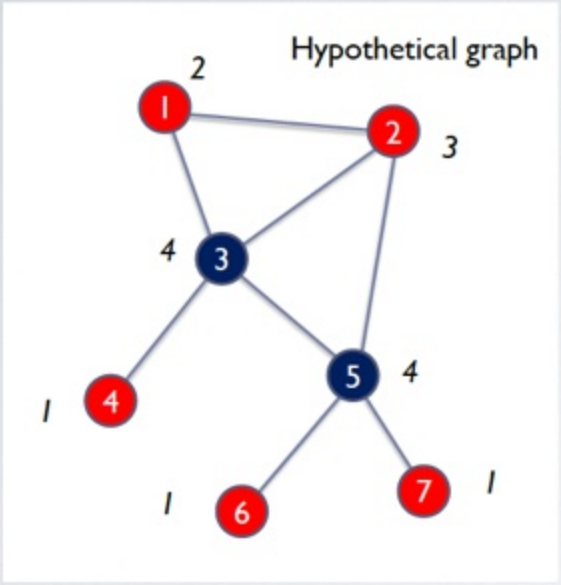




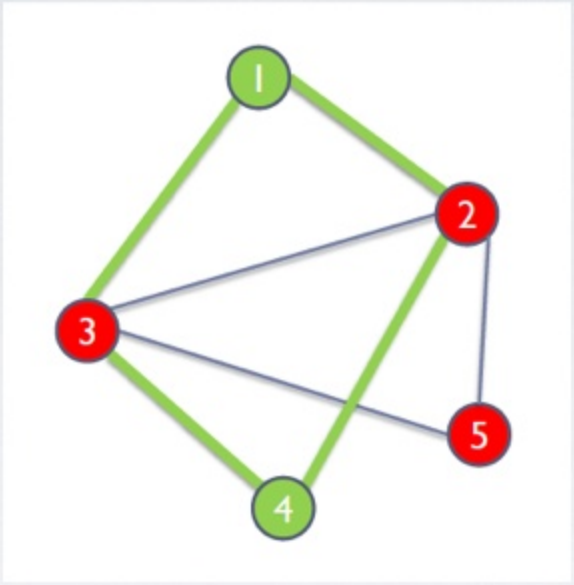


Different Network, Same Graph

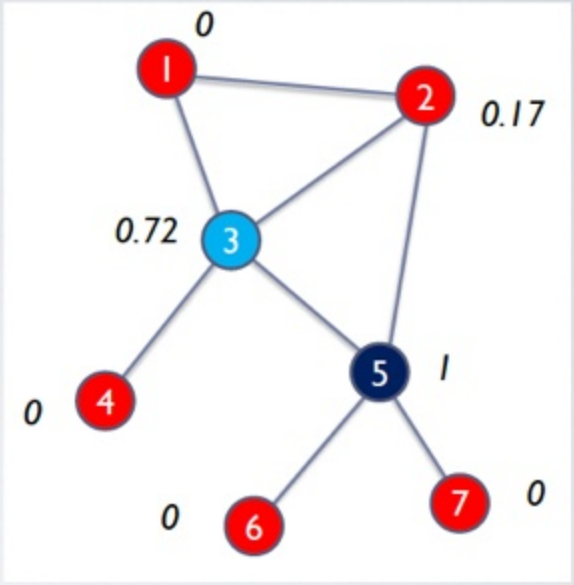
METRIK CENTRALITY



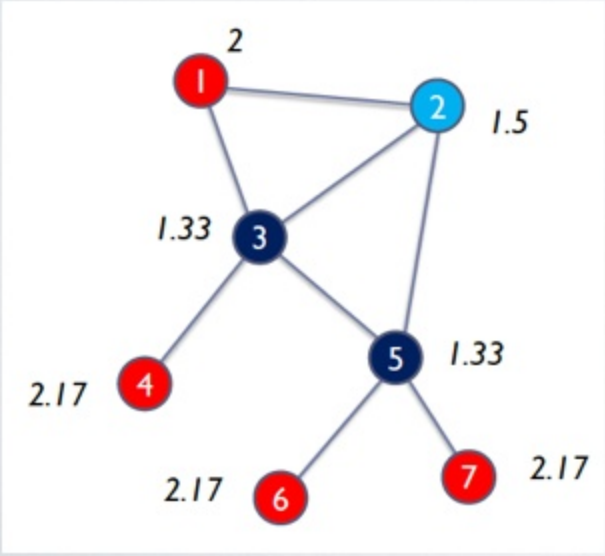
degree centrality



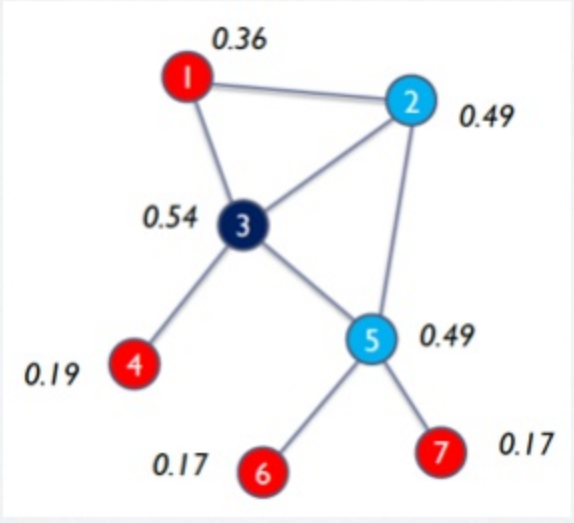
shortest path



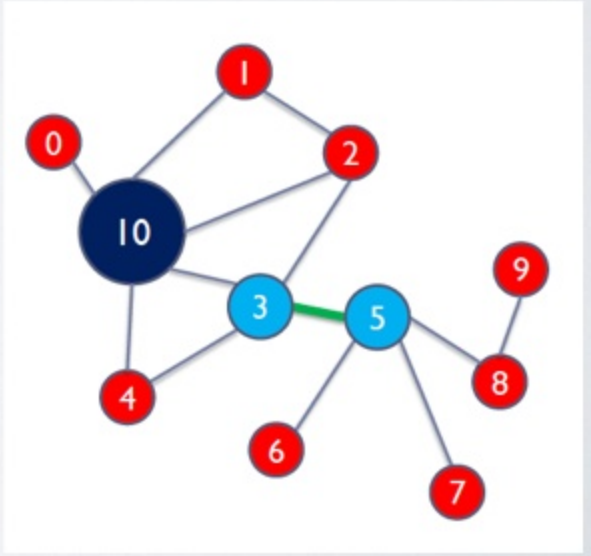
betweenness centrality



closeness centrality

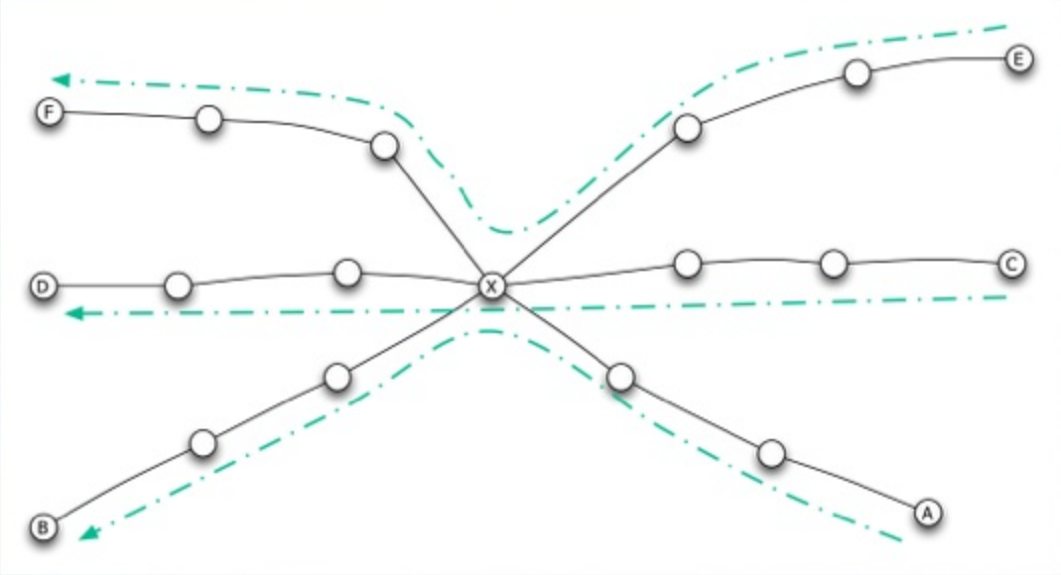


eigenvector centrality



set of key players

METRIK CENTRALITY



betweenness centrality

banyaknya **jalur terpendek** antar pasangan semua titik di jaringan, yang melewati satu titik yang diukur

closeness centrality

jarak titik yang diukur terhadap semua titik yang ada dalam jaringan