

GUG 2nd Seminar@ Sep 23, 2023

고차 상호작용 기반의 하이퍼그래프 모델링

곽상환 PhD

2023.09.23.Sat

Overview

1. 그래프와 네트워크에 관하여
2. 네트워크 과학이란 무엇인가?
3. 하이퍼 그래프란 무엇인가?
4. 네트워크 과학에서의 하이퍼 그래프
5. 데이터베이스에서의 하이퍼 그래프

Graph: Introduction

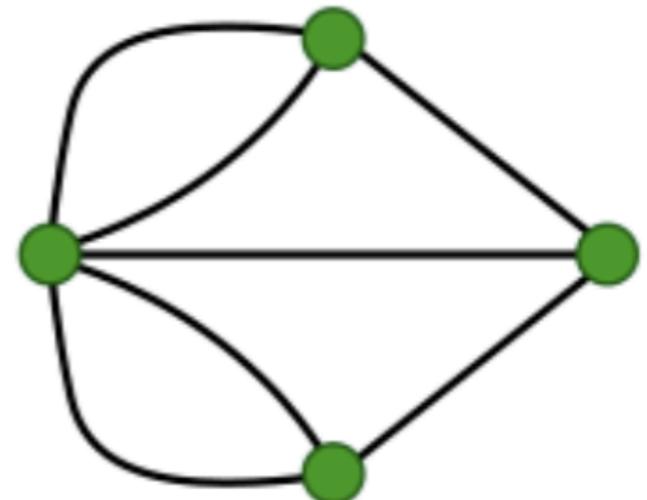
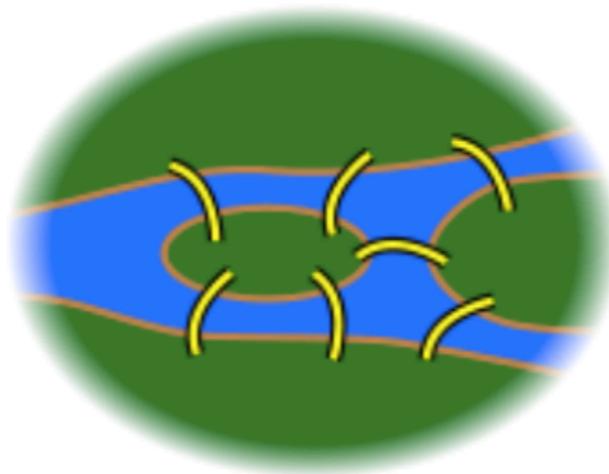
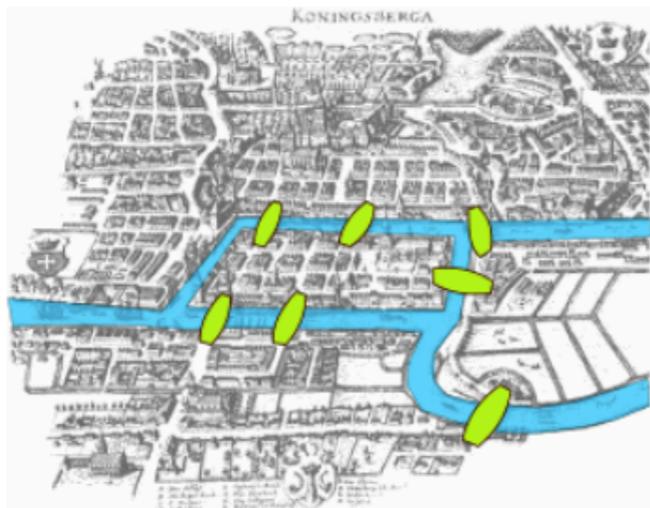


- Begun in 1735.
- Leonhard Euler's paper "Seven bridge of konigsberg"

“일곱개의 다리들을 한 번씩만 차례로 모두 건널 수 있겠는가?”

“없다.

**모든 점이 짝수개의 선을 가지거나
단 두개의 점만이 홀수개의 선을 가져야 한다.”**



Graph: Introduction

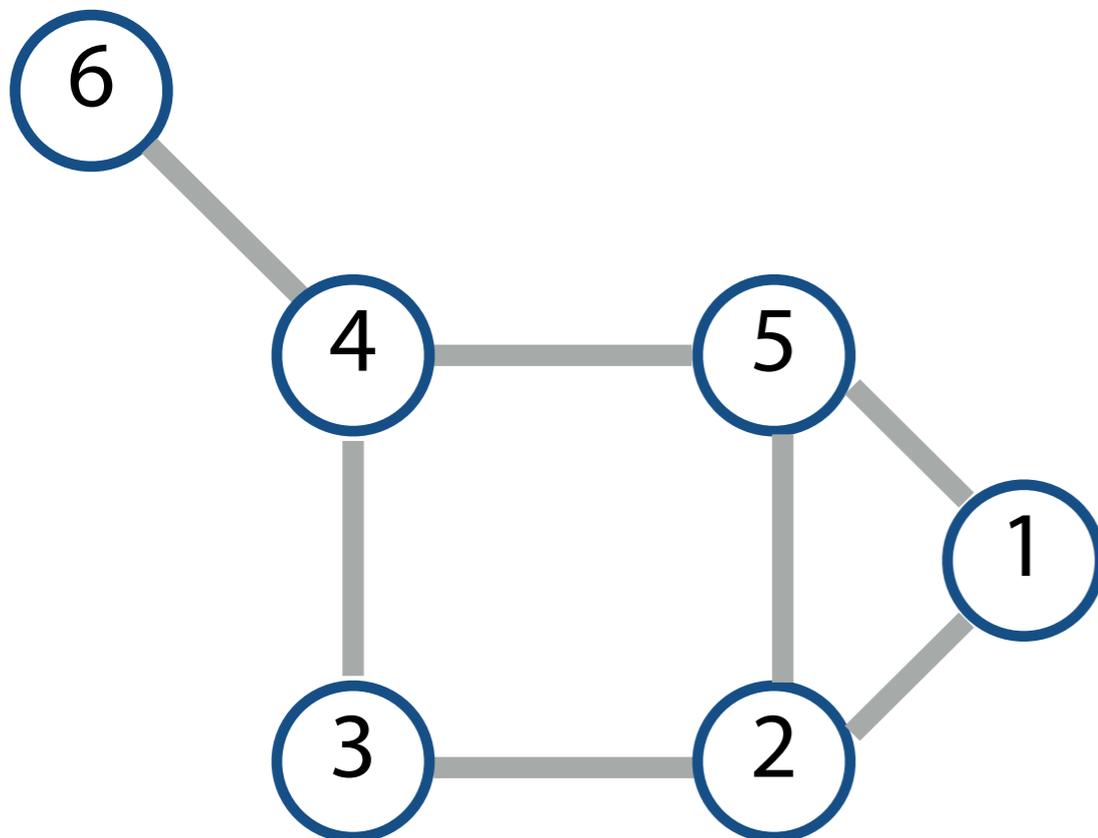
- A graph is a collection of nodes and edges.

Denoted by $G(V, E)$.

V =nodes(**vertices, sites, points..**)

E =links(**edges, bonds, arcs..**) between two pairs of nodes.

Graph size parameter $N = \sum |V|, M = \sum |E|$.



$$N=6, M=7$$

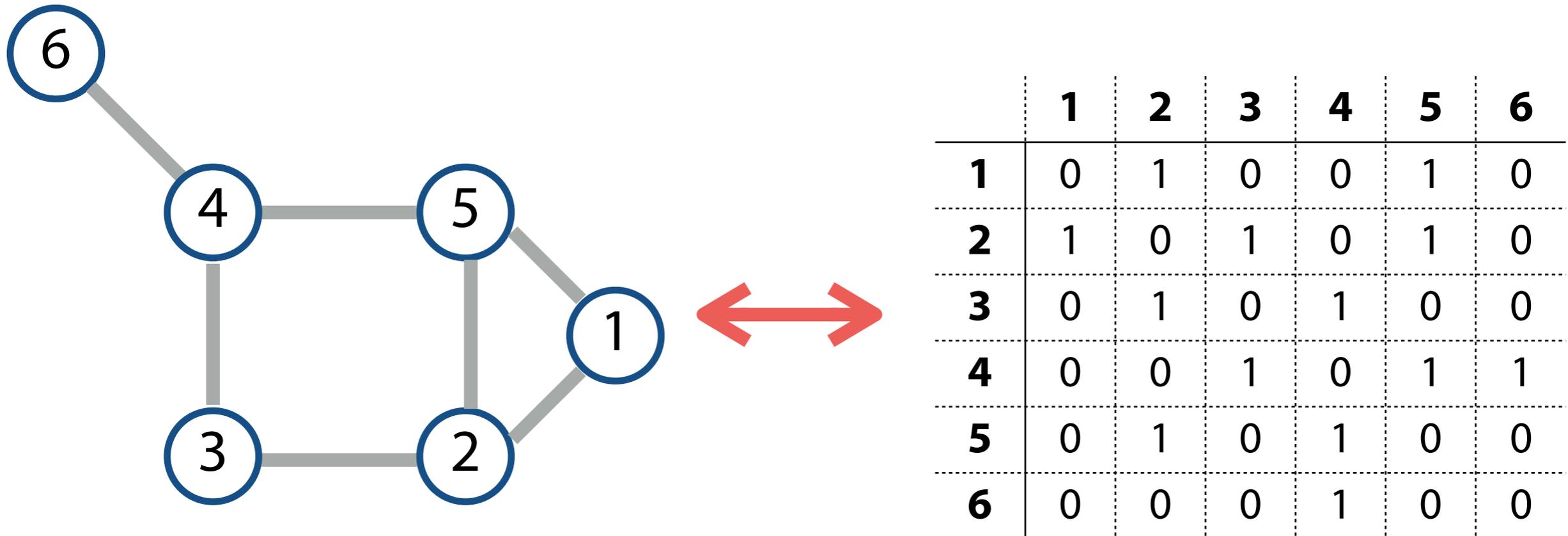
Vertices $(V) = \{1, 2, 3, 4, 5, 6\}$

Edge $(E) = \{1, 2\} \{1, 5\} \{2, 3\} \{2, 5\}$

$\{3, 4\} \{4, 5\} \{4, 6\}$

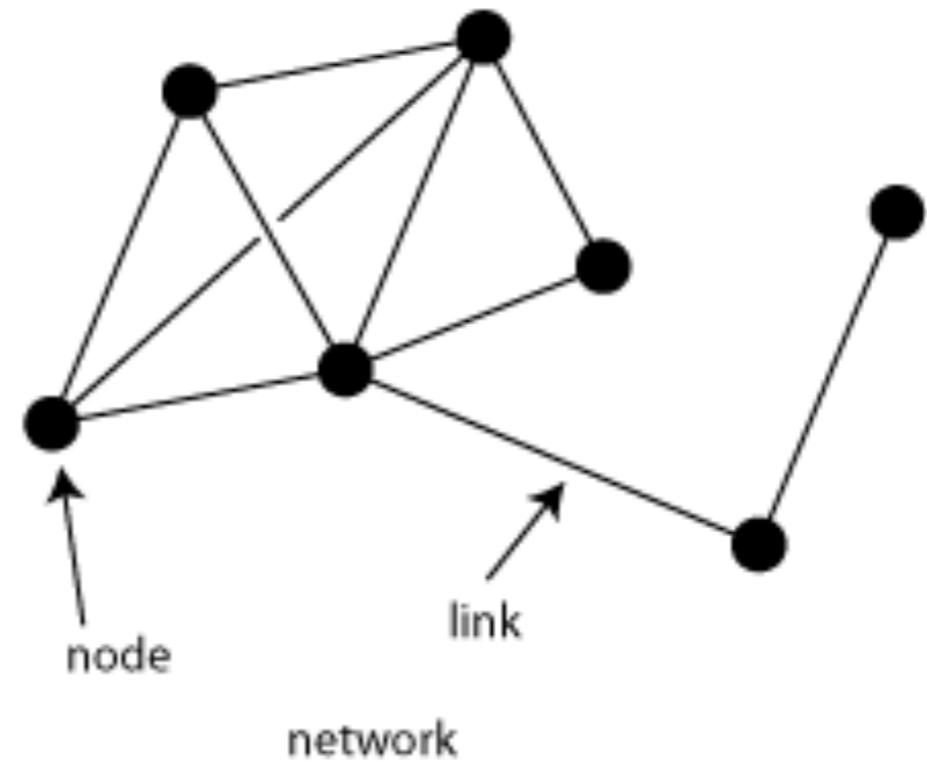
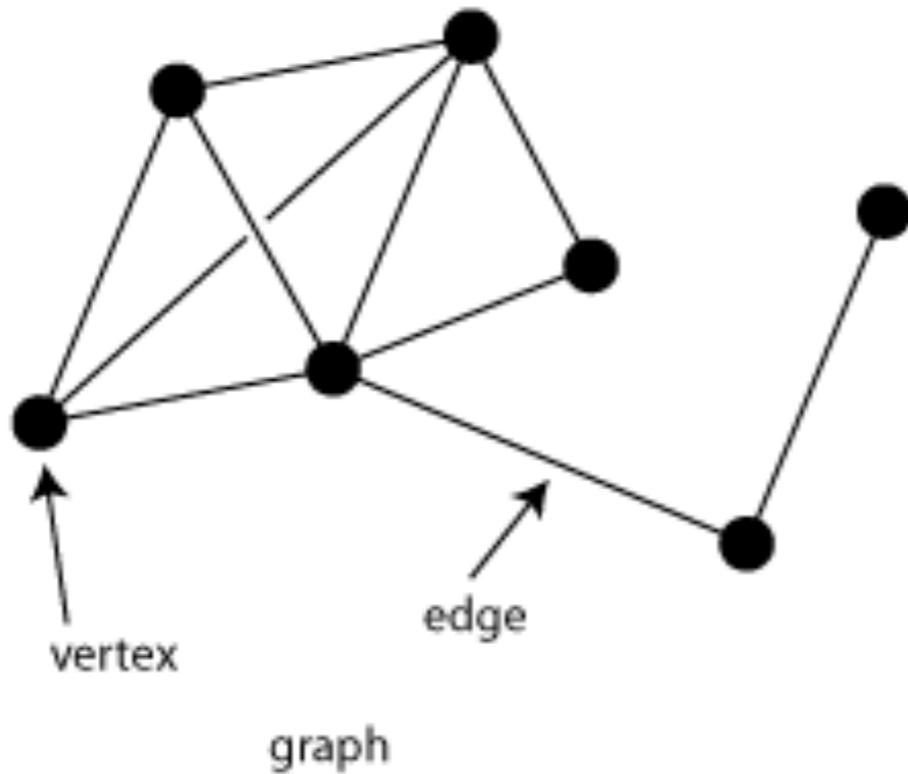
Ex) Nearest Neighbor $V(4) = \{3, 5, 6\}$

Simple Graph; Representation



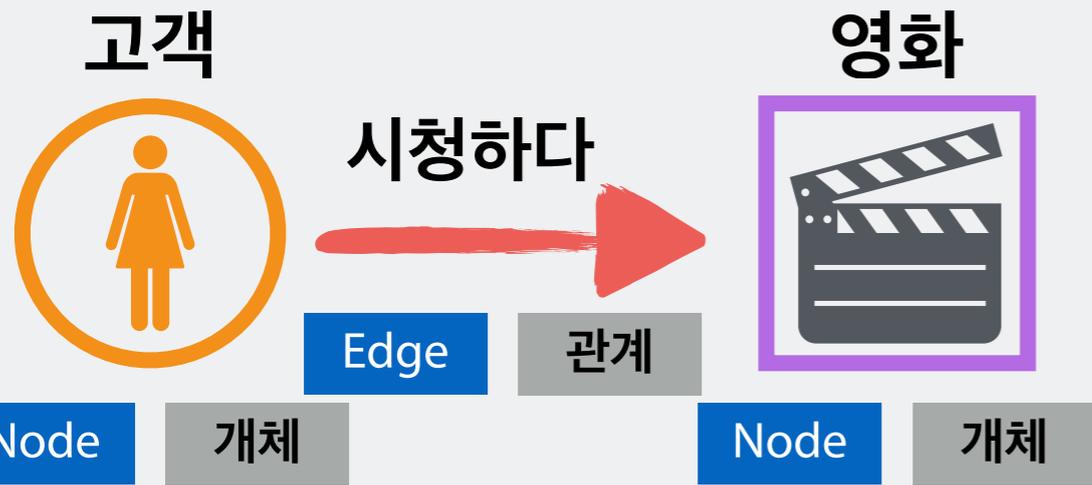
- n -by- n matrix with $A_{uv} = 1$ if (u, v) is an edge.
- Diagonal entries are self-loops.
- Symmetric matrix for undirected graph

Graph? Network?



Network Science	Node	Link	Cluster
Graph Theory	Vertex	Edge	Component
Statistical Physics	Site	Bond	Component

추천 시스템 활용 *using Gprah DataBase*



▶ **고객 흥미 분석:** 이 시나리오에서는 'travel' 및 'comedy'와 같은 여러 종류의 **영화 장르** 또는 관련된 정보를 검색하여 **고객의 흥미를 분석**.

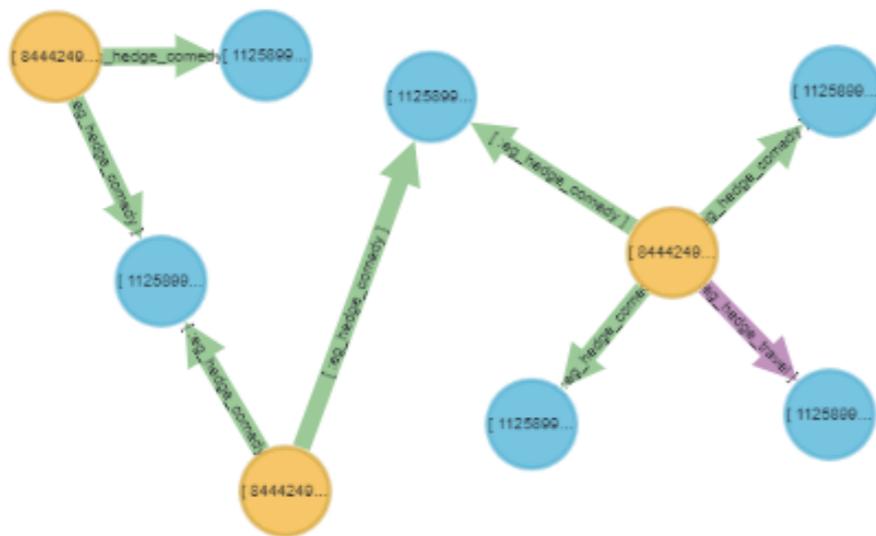
Query

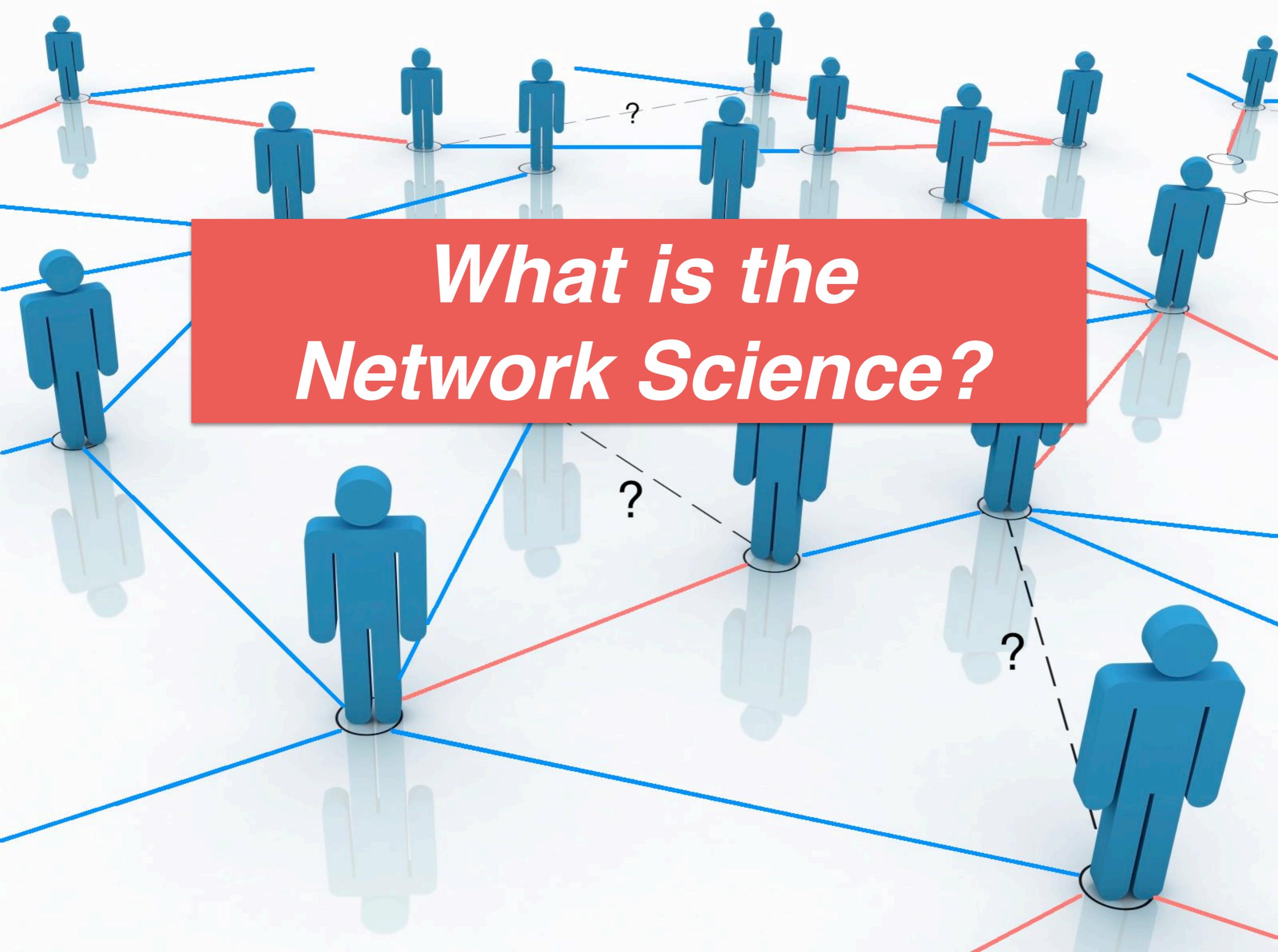
```
SELECT * FROM cypher('hgdb_dvd_ness', $$
MATCH p=(d:vt_customer)-[r]->() WHERE type(r) =
'eg_hedge_travel' OR type(r) = 'eg_hedge_comedy'
RETURN p LIMIT 10 $$) AS result(p agtype);
```

결과

어떤 고객이 코미디, 여행 장르의 영화에 흥미를 보였고 그 영화를 보지 않았지만 **다른 고객에게 영화를 추천**.

고객들에게 관심사에 맞는 영화를 제안하는 데 활용.





***What is the
Network Science?***

Definition of Network Science

Albert-László Barabási

"네트워크는 연결된 객체 또는 노드들로 구성된 시스템으로, 이러한 노드들 간의 상호작용이나 연결 관계를 연구하고 모델링하는 과학"

"Network science is the study of systems composed of interconnected objects or nodes, focusing on understanding and modeling the interactions or connections between these nodes"

Duncan J. Watts

"네트워크 과학은 복잡한 시스템의 구성 요소 간의 상호작용을 연구하는 학문으로, 그 중에서도 그 구조와 특성을 중점적으로 다룬다"

"Network science is a discipline that investigates the interactions among components of complex systems, with a particular emphasis on the structure and characteristics of these interactions"

Mark Newman

"네트워크 과학은 다양한 사회, 생물학, 기술 분야에서 발생하는 복잡한 관계를 분석하고 모델링하기 위한 수학적 도구와 기법을 개발하며, 이를 통해 시스템의 구조와 동작을 이해하는 학문"

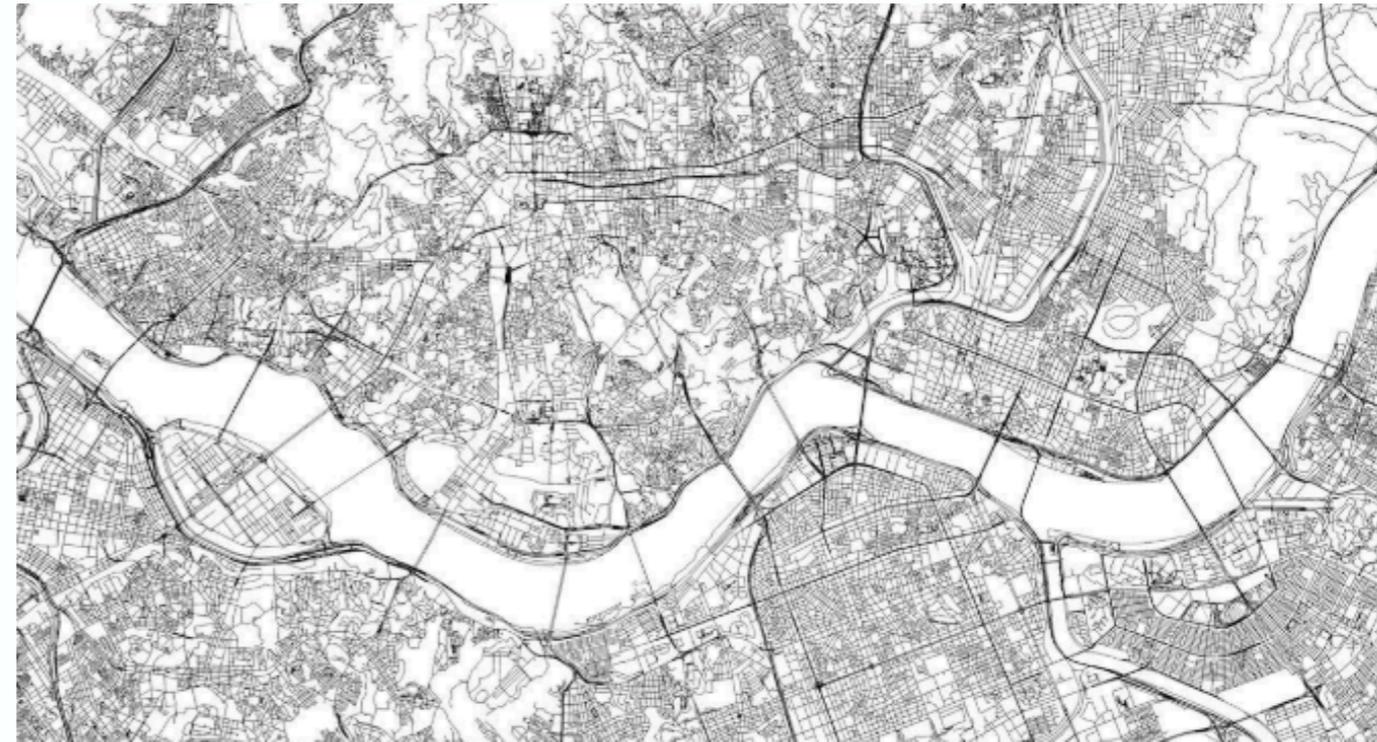
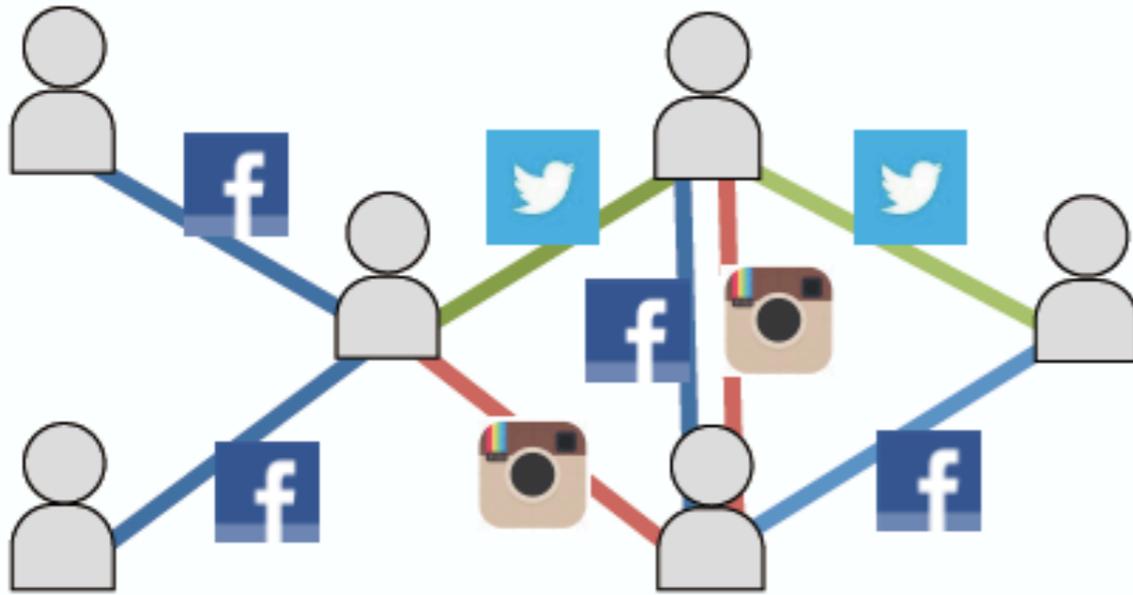
"Network science is a field that develops mathematical tools and techniques for analyzing and modeling complex relationships that arise in various domains such as social, biological, and technological systems. It aims to understand the structure and behavior of systems"

“

연결된 객체나 노드들로 이루어진 시스템의 상호작용, 구조, 특성을 연구하고 모델링하는 학문

”

Networks and Data



The global avalanche network

in "Impact of the topology of global macroeconomic network on the spreading of economic crises"

by K.-M. Lee, J.-S. Yang, G. Kim, J. Lee, K.-I. Goh, and I.-M. Kim © 2011

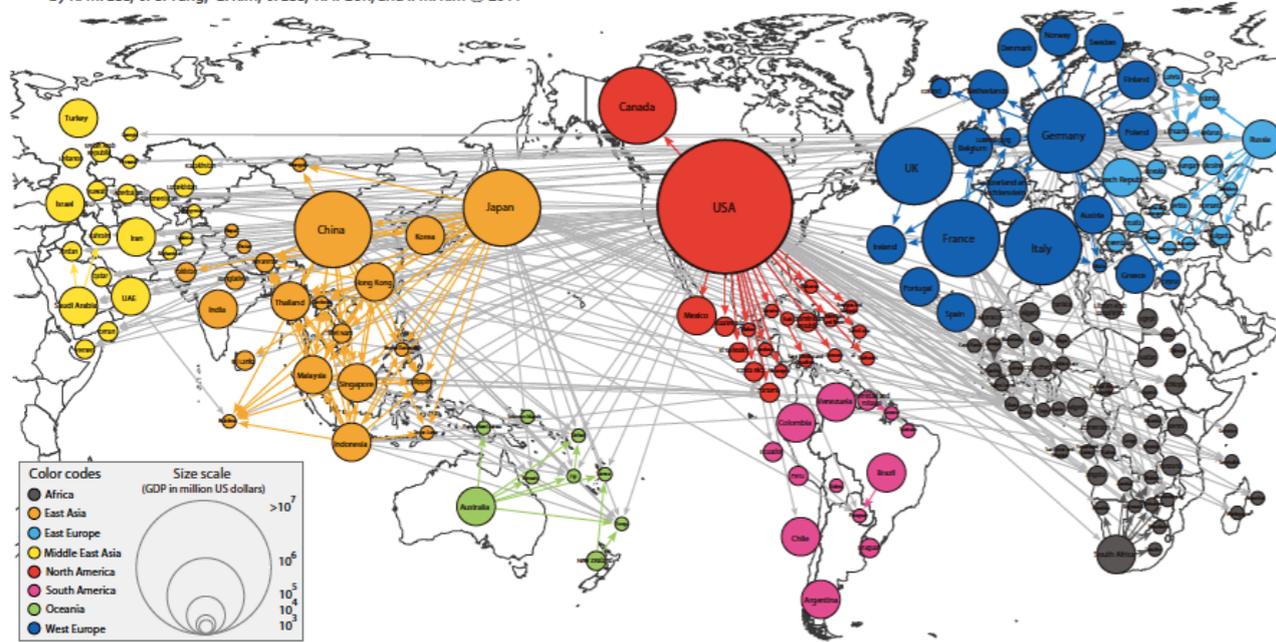
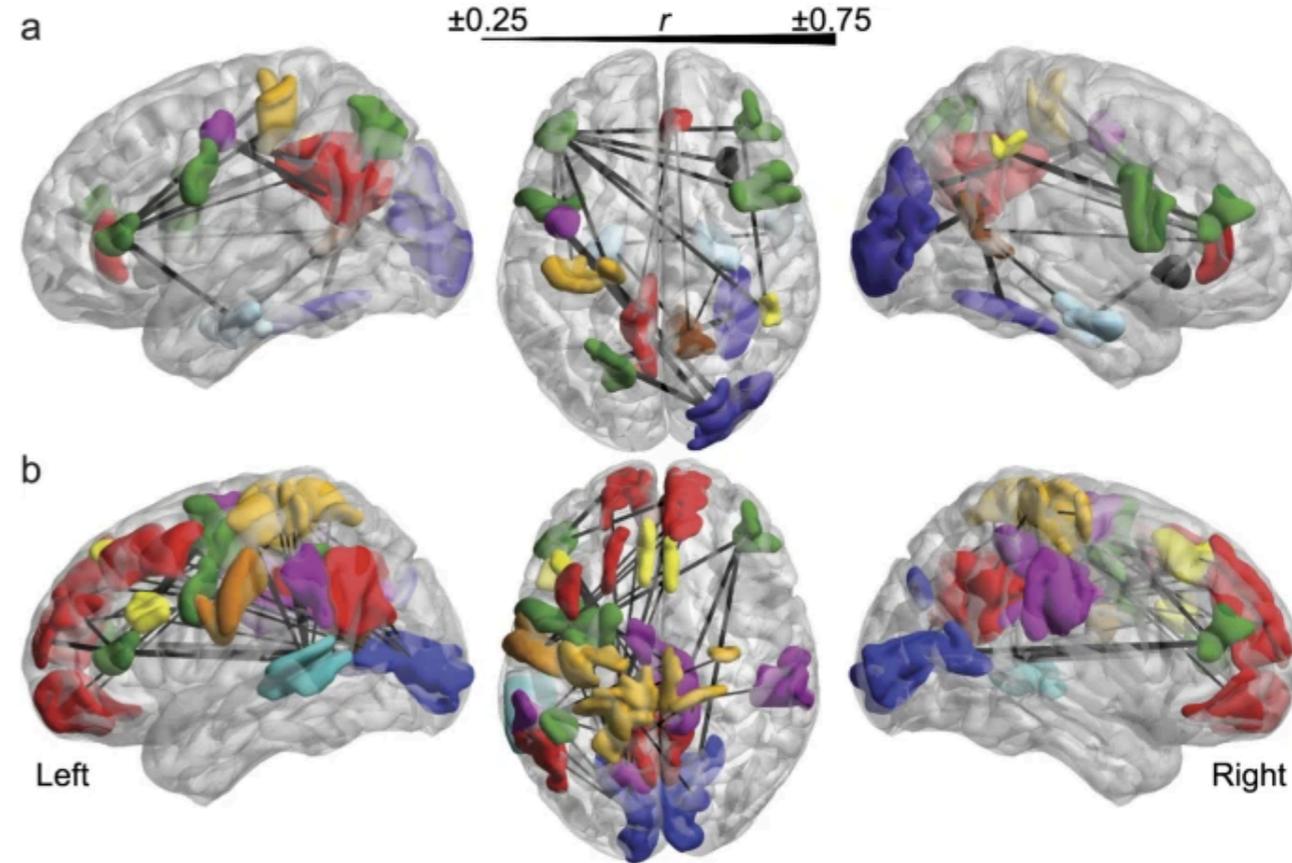
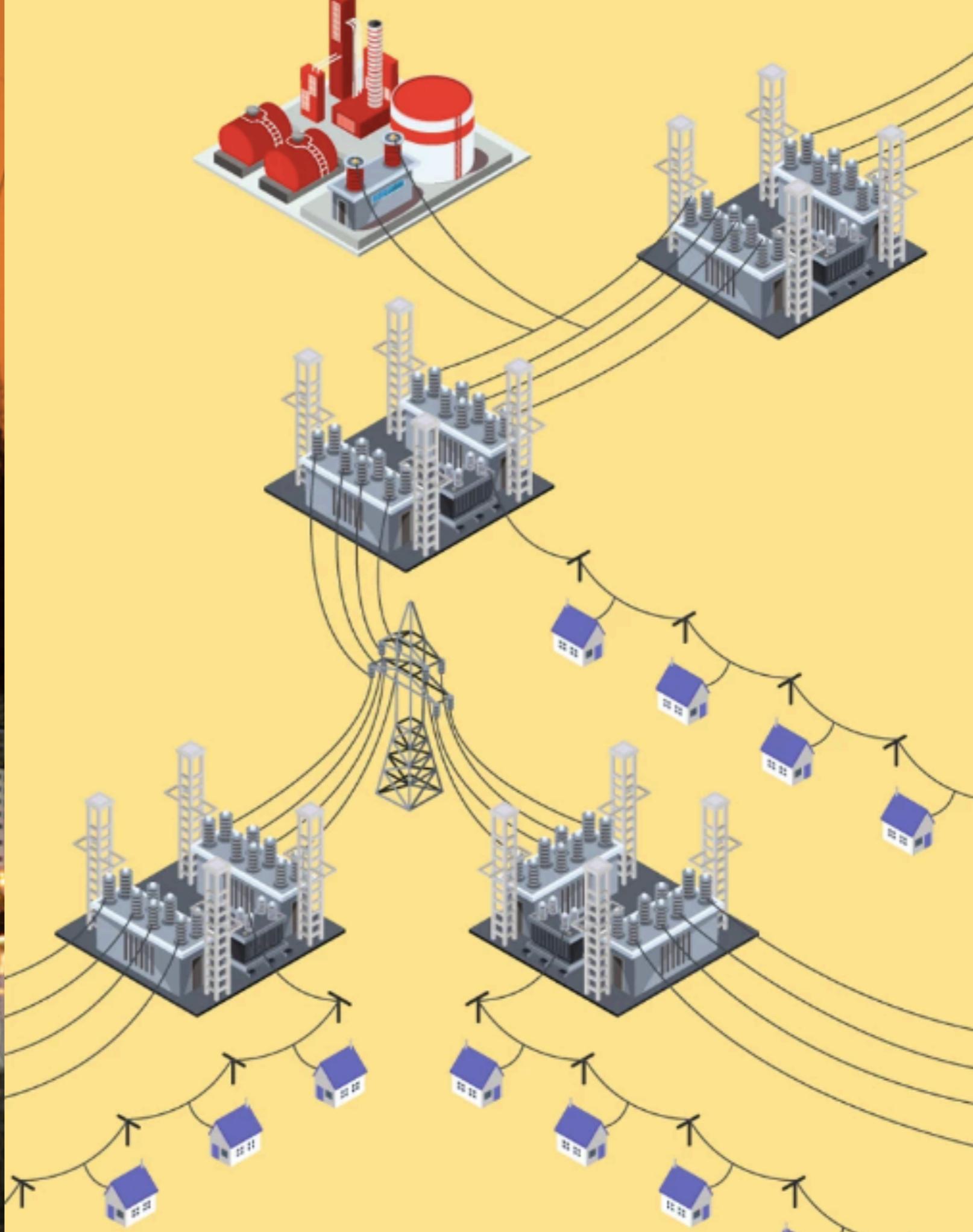
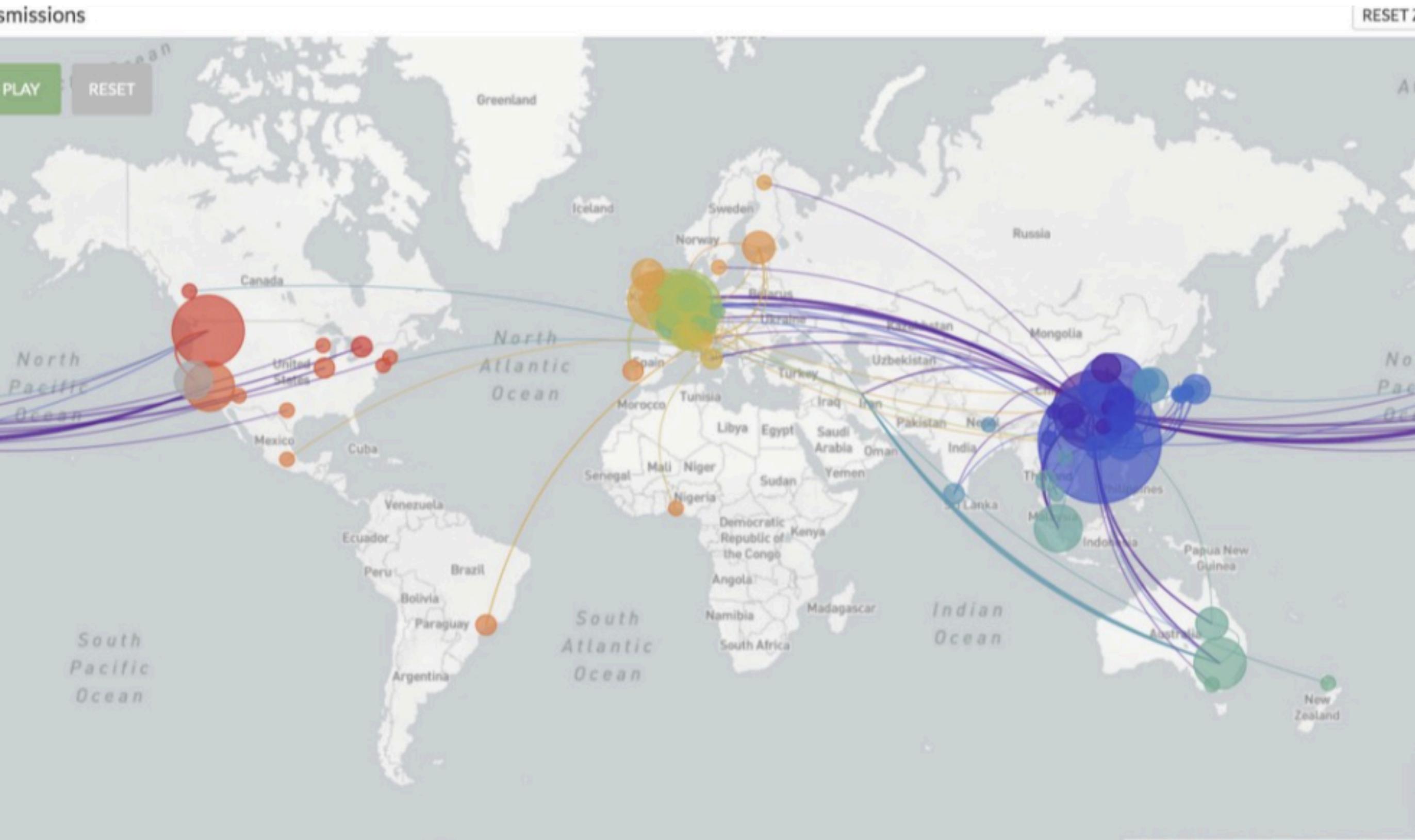


Figure 5A. The avalanche network for the world economic network at 10:17 overlaid with the world map. The countries are color-coded according to the continental association and the size of a node follows its GDP (see legend). An arrow from A to B is placed if the country A makes B collapse in the crisis spreading model. The arrow connecting countries within the same continent is colored following the same color code, and otherwise it is colored grey.





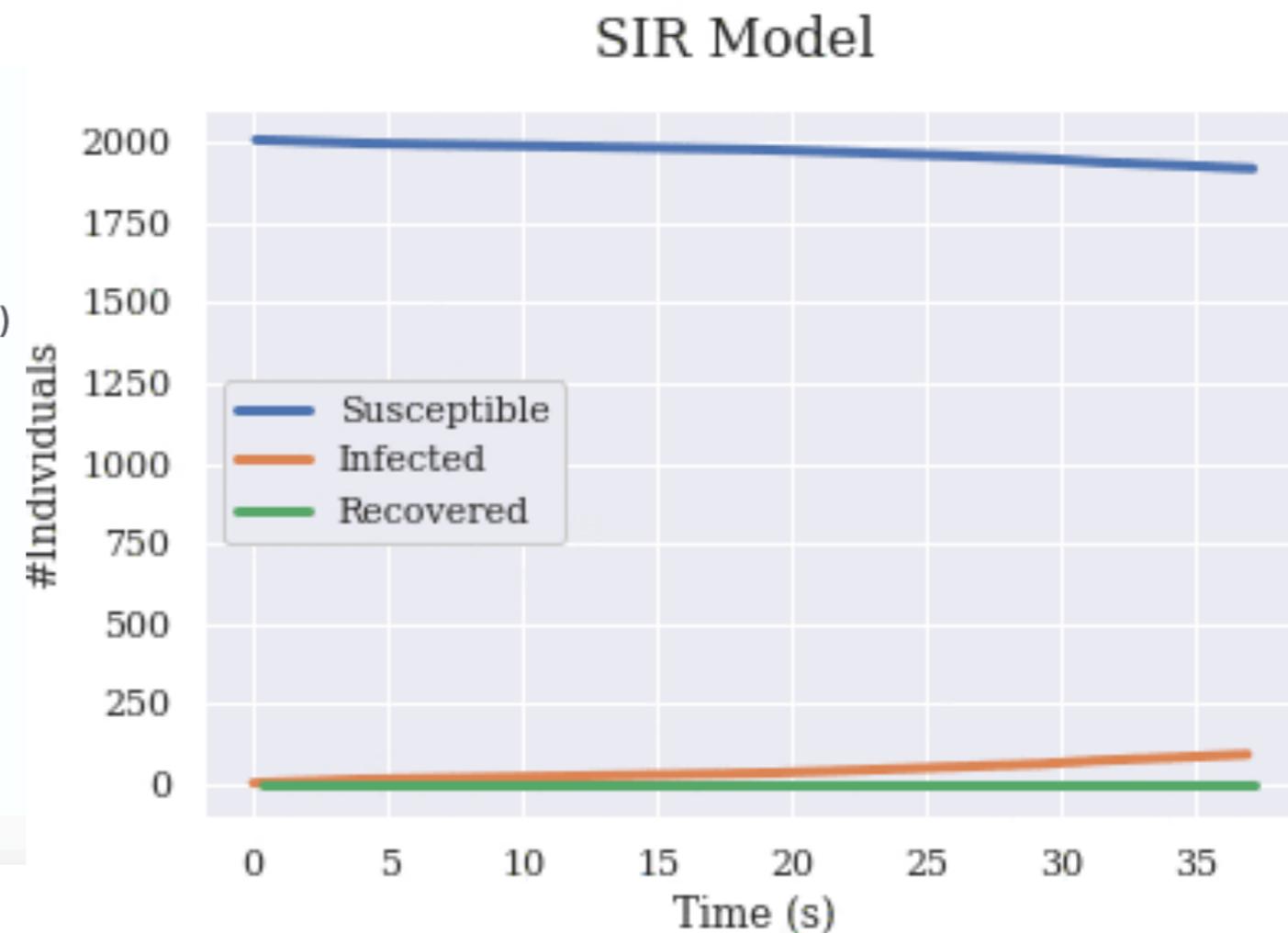
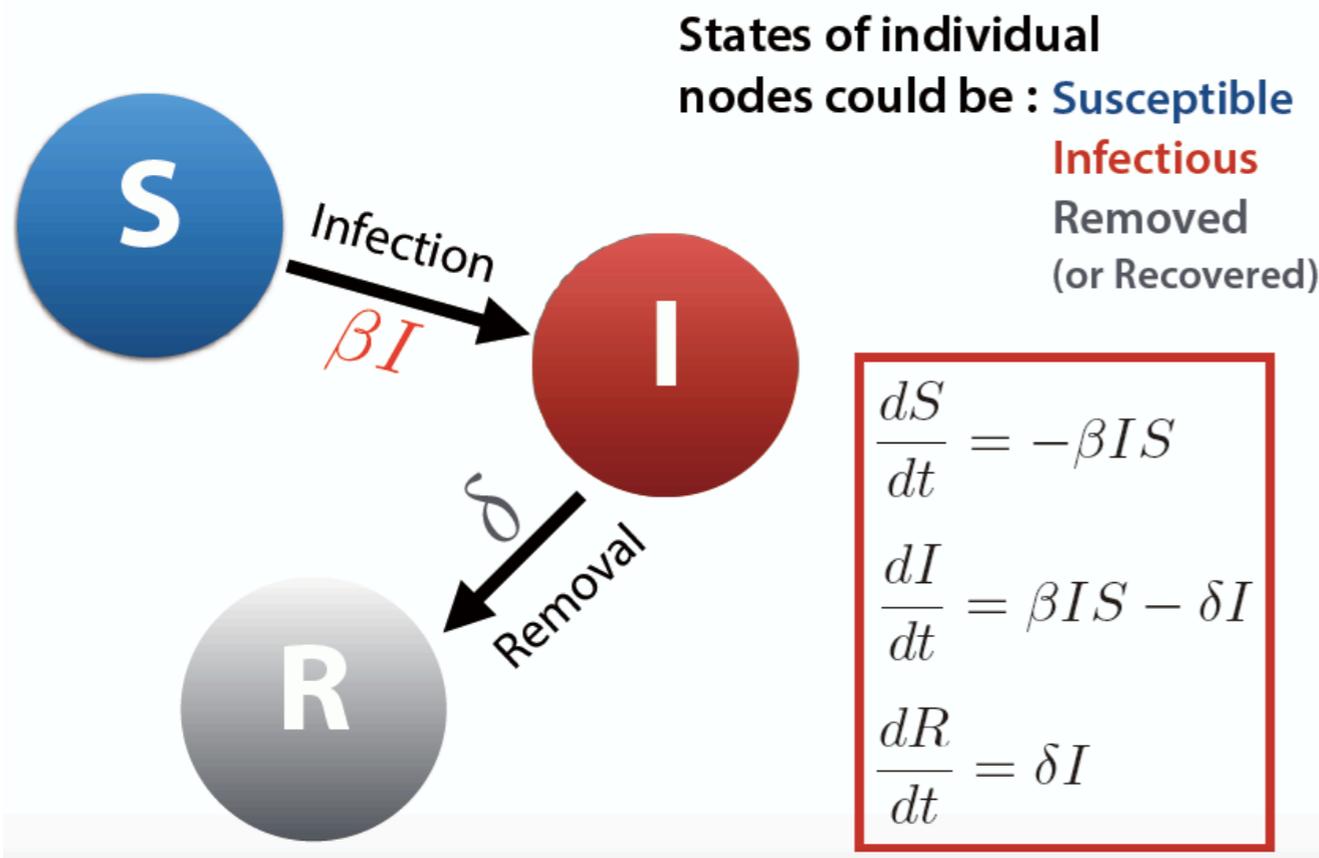
• *Example of the virus : COVID-19*



Network dynamics; Disease spreading

Graph Scientist, What can we do??

- I. Recognition of disease characteristic.
- II. Find hub patients and effective quarantine.
- III. Data collection and analysis for modeling.



Network dynamics; Disease spreading

Eq. 2 $\frac{dx}{dt} = (\beta s - \gamma)x$
 $= \gamma(R_0 s - 1)x.$

→ $R_0 s \geq 1$

→ $R_0(1 - r) \gtrsim 1$

Herd immunity threshold:

$$r_H = 1 - \frac{1}{R_0}.$$

감염자 생산지수

(감염자 한 사람이 몇 명을 전염시키는지에 대한 지표)

$$R_0 = \frac{\beta}{\gamma}$$

유명한 질병의 R0 값^[1]

질병	전염경로	R ₀
홍역	공기매개	12-18
디프테리아	타액	6-7
SARS-CoV-2 델타 변이	비말	5-9
천연두	비말	5-7
소아마비	대변-구강 경로	5-7
풍진	비말	5-7
볼거리	비말	4-7
백일해	비말	5.5 ^[2]
코로나바이러스감염증-19	비말	3-5 ^{[3][4]}
HIV/에이즈	성적 접촉	2-5
사스	비말	2-5 ^[5]
인플루엔자 (1918년 범유행 strain)	비말	2-3 ^[6]
에볼라 (2014년 에볼라 유행병)	체액	1.5-2.5 ^[7]

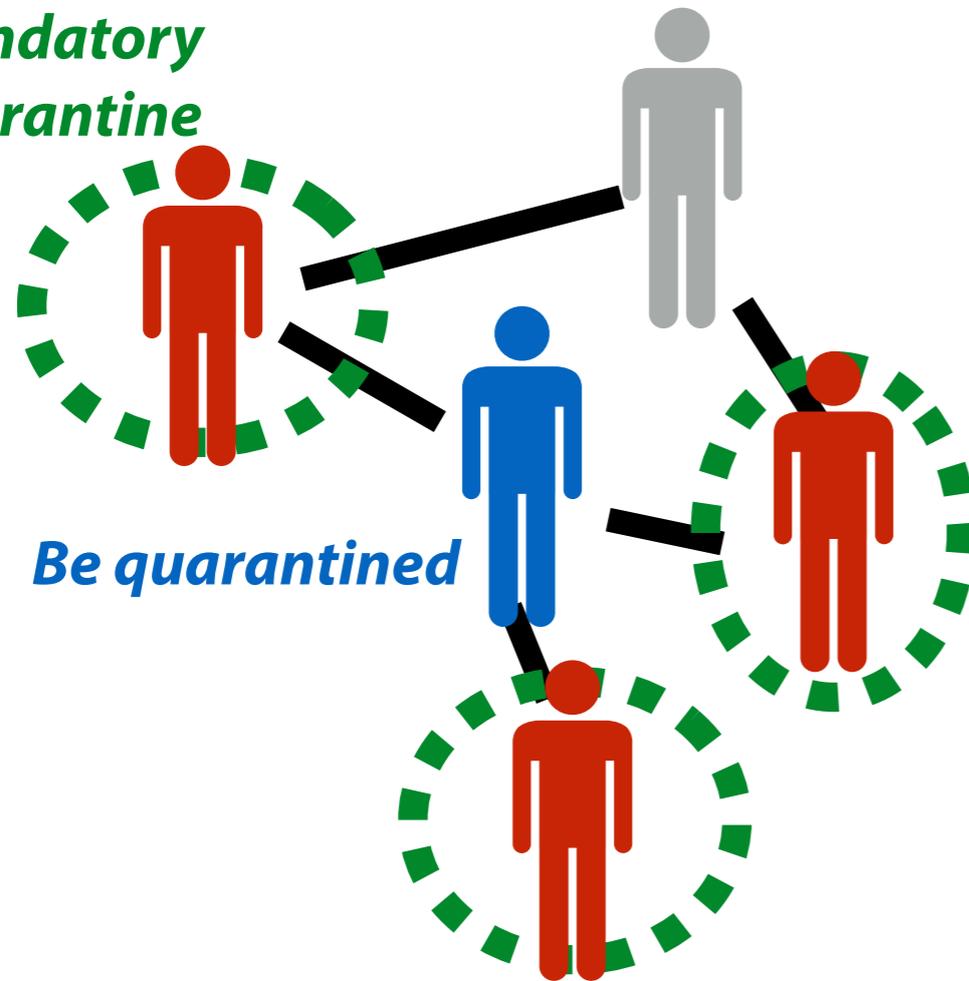
Disease spreading on social networks

- **Assumed that, disease spreads through the social networks.**

=> Sintayehu, Dejene W., et al. Preventive veterinary medicine 137 (2017): 36-42.

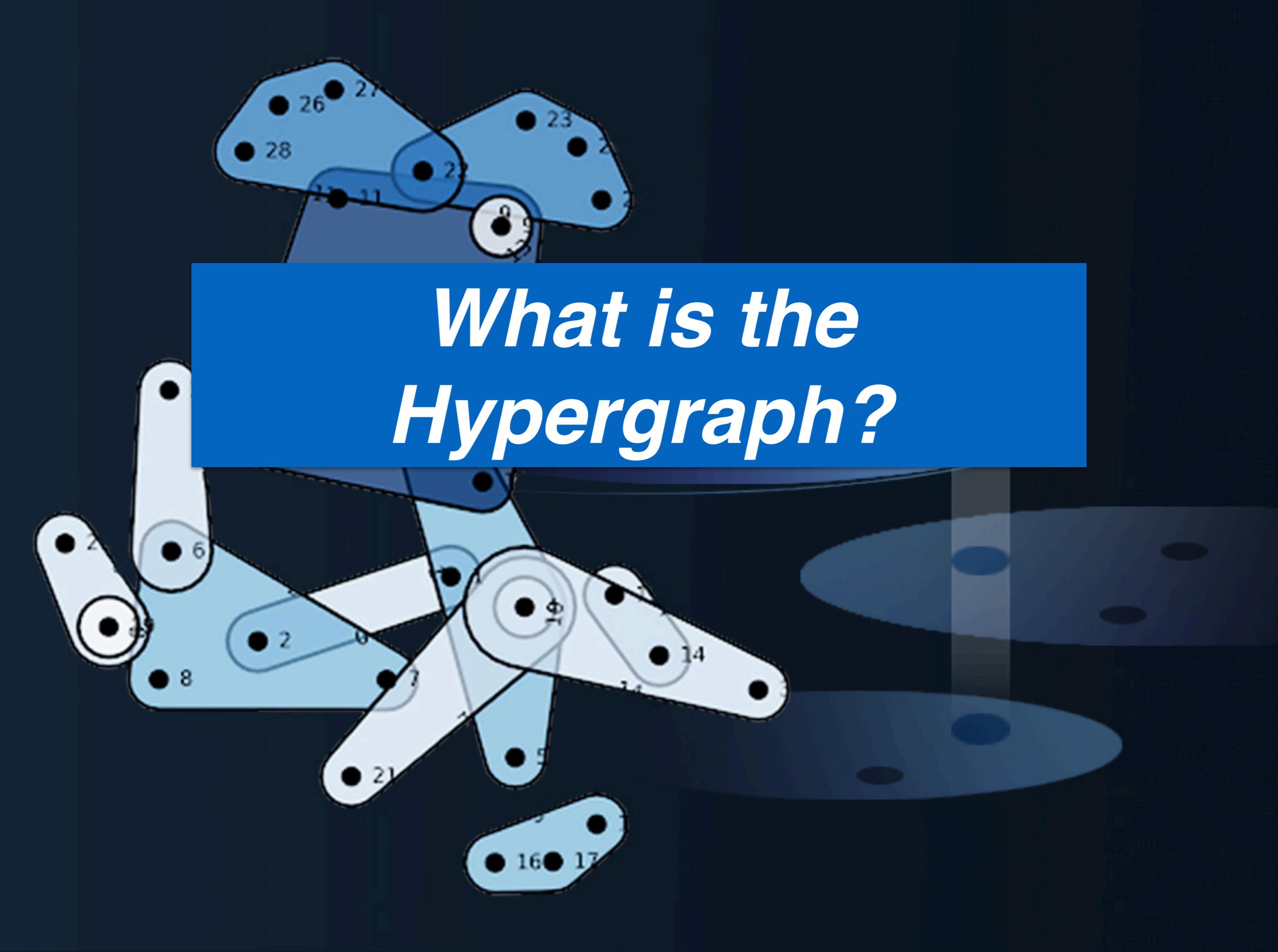
=> Strano, Emanuele, et al. Scientific reports 8.1 (2018): 4744.

Mandatory quarantine



In percolation view,

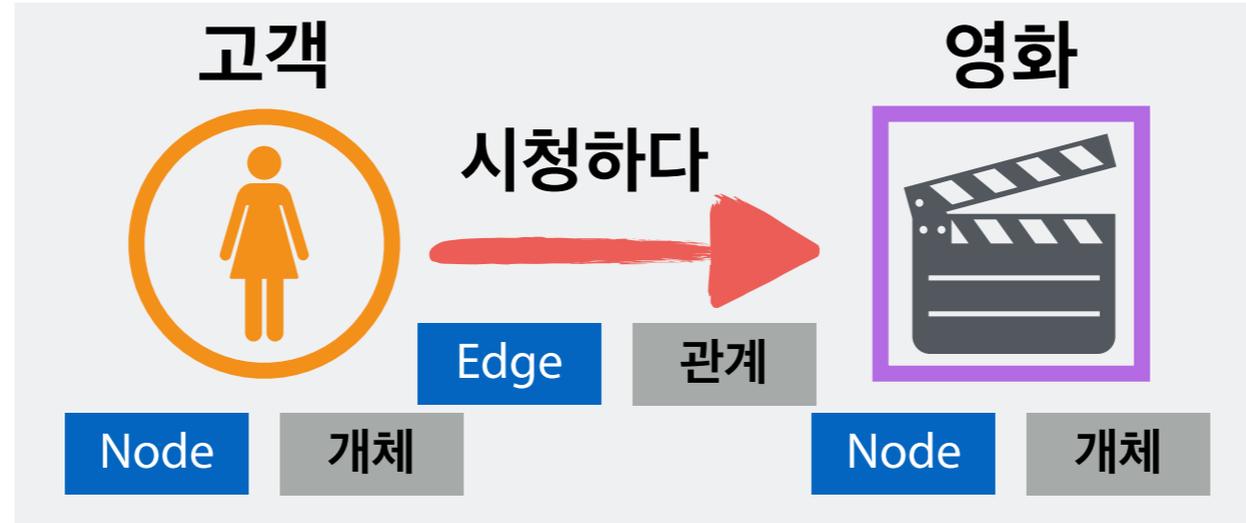
- **Mandatory quarantine**
>> **Random occupied cluster (or nodes)**
- **Be quarantined**
>> **Exclaves cluster (or nodes)**



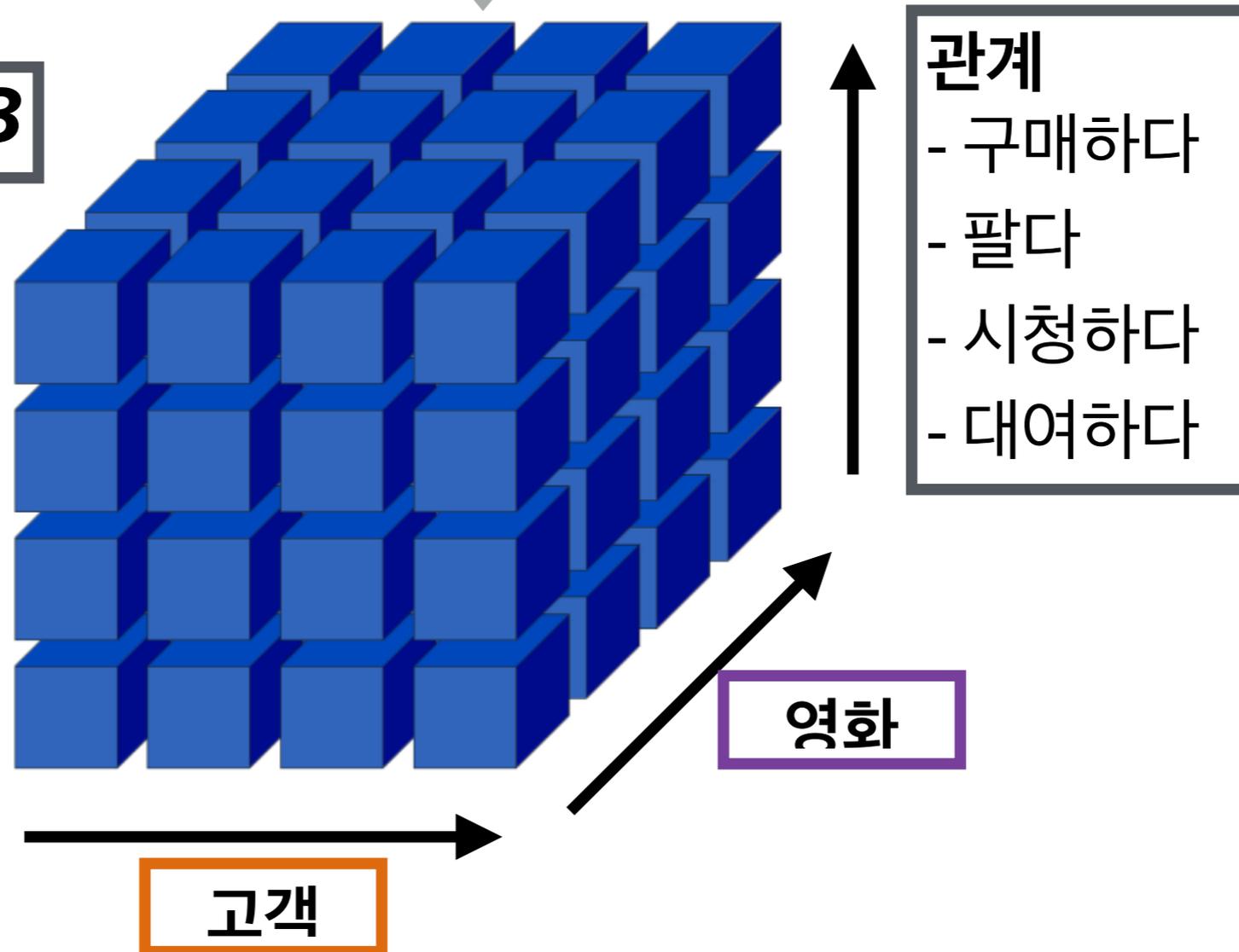
*What is the
Hypergraph?*

Motivation of Hypergraph

Graph DB

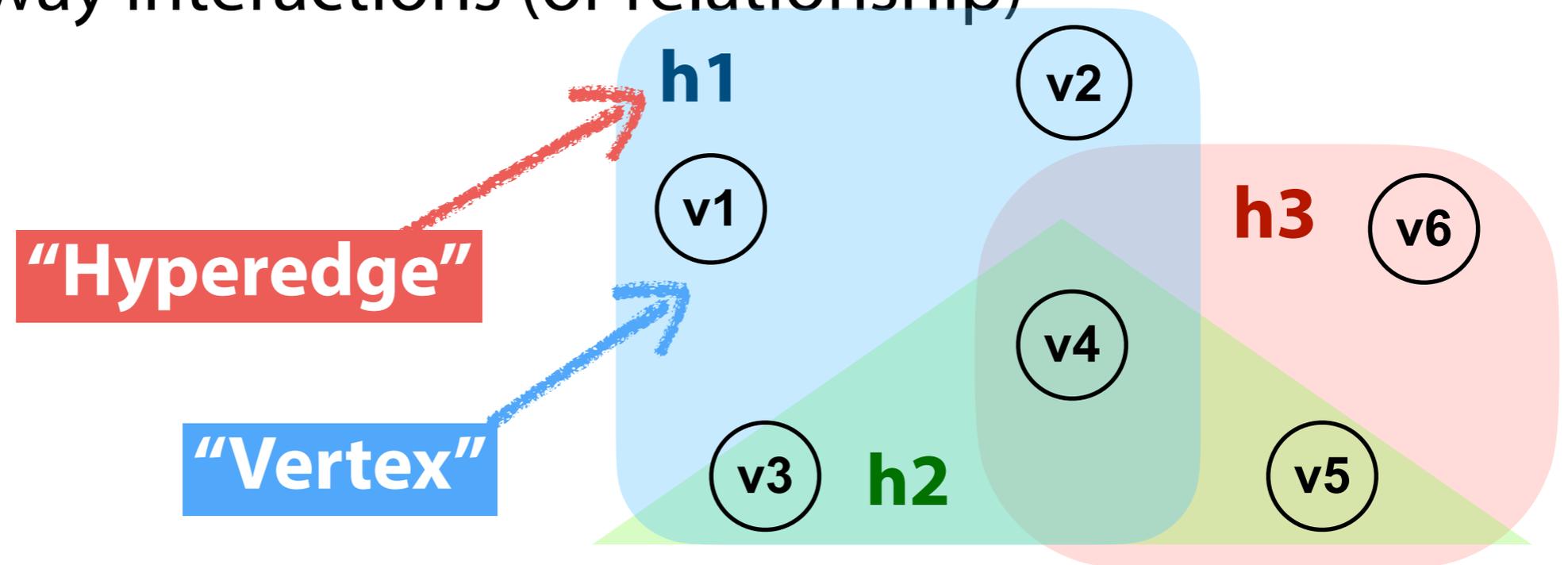


Hypergraph DB



Definition of Hypergraphs

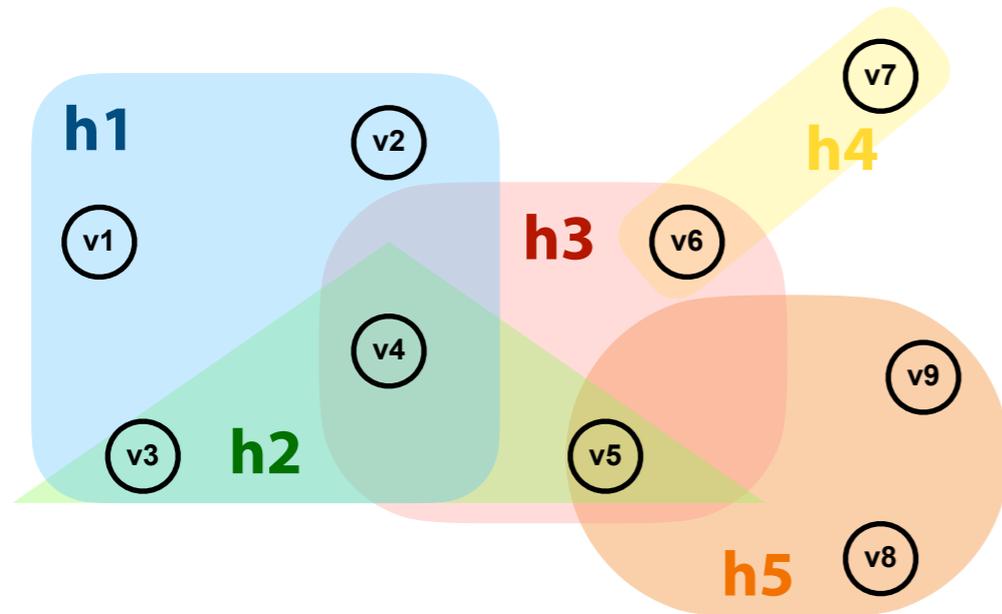
- ▶ **Hypergraphs** provide a mathematical model of data focused on multi-way interactions
 - ▶ To ask certain kinds of questions
 - ▶ Connectivity of entities
 - ▶ Clustering structure
 - ▶ To model certain kinds of interactions
 - ▶ Multi-way interactions (or relationship)



Definition of Hypergraphs

Hypergraphs Representation

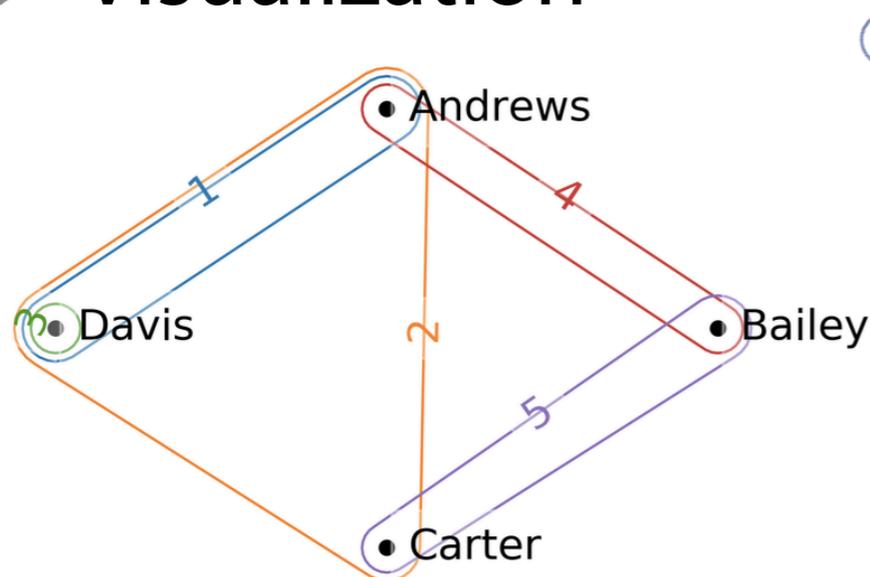
Matrix or Complete Set Notation (Hyperedge list)



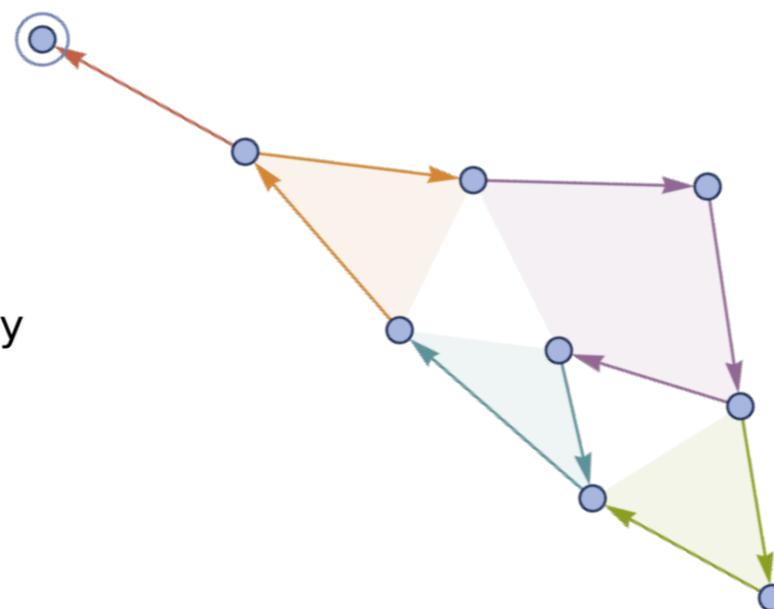
	v1	v2	v3	v4	v5	v6	v7	v8	v9
h1	1	1	1	1	0	0	0	0	0
h2	0	0	1	1	1	0	0	0	0
h3	0	0	0	1	1	1	0	0	0
h4	0	0	0	0	0	1	1	0	0
h5	0	0	0	0	1	0	0	1	1

Hypergraphs Representation

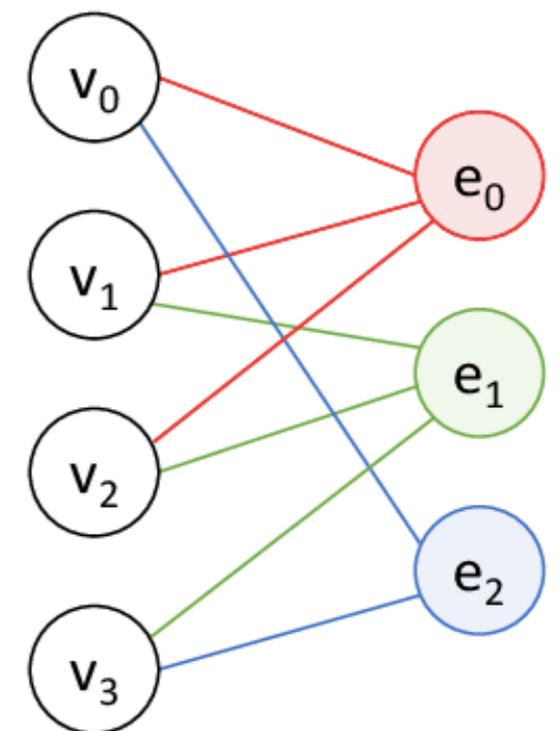
Visualization



Euler diagram



Simplicial diagram



Biprtite diagram

Definition of Hypergraphs

Hypergraphs properties

1. Degree distribution(하이퍼 엣지의 분포)
2. Edge size distribution
3. s-Walk, s-Path, s-Diameter
 - 3.1.s-Walk: Sequential edges, allowing repeated edge traversal, indicating movement along connected vertices in a hypergraph.
 - 3.2.s-Path: Permutation of edges connecting distinct vertices in a hypergraph, showing their traversal order.
 - 3.3.s-Diameter: Maximum length among the shortest s-paths between all vertex pairs in a hypergraph, revealing the furthest distance.
4. s-Connected components
 - 4.1.Subset of connected vertices by s-paths, signifying a cohesive substructure.
5. s-Centrality(중심성)
 - 5.1.Measures vertex importance based on s-path participation.
6. Clustering coefficient(클러스터링 계수)
7. Find the loop or triangular

Application to Hypergraphs: Network Science

Project: World trade network

The global avalanche network

in "Impact of the topology of global macroeconomic network on the spreading of economic crises"

by K.-M. Lee, J.-S. Yang, G. Kim, J. Lee, K.-I. Goh, and I.-M. Kim © 2011

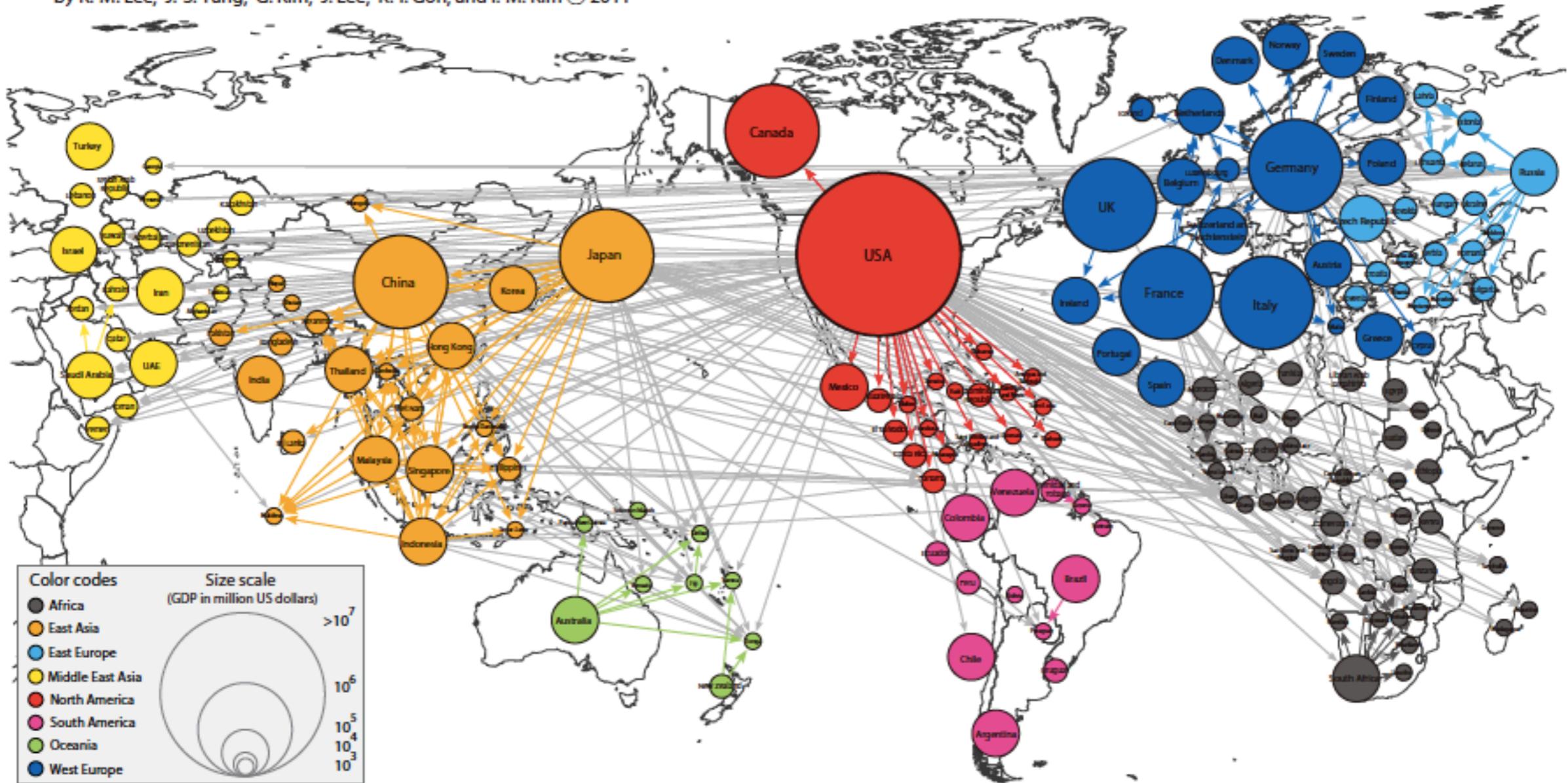


Figure S4. The avalanche network for the world economic network at $t=7$, overlaid with the world map. The countries are color-coded according to the continental association and the size of a node follows its GDP (see Legend). An arrow from A to B is placed if the country A makes B collapse in the crisis spreading model. The arrow connecting countries within the same continent is colored following the same color codes, and otherwise it is colored gray.

Project: World trade network

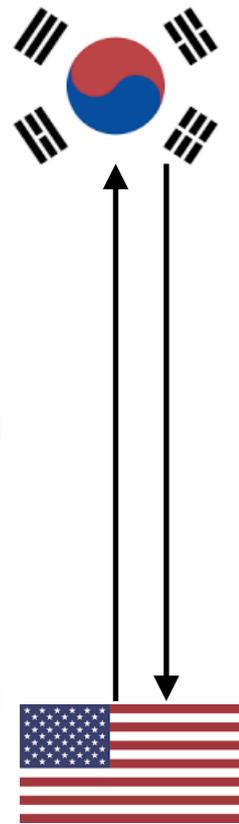
Country1, Country2, Product #, Volume

1	135040	117100	0370	10
2	135040	117100	0459	11
3	135040	117100	0483	3
4	135040	117100	0541	1
5	135040	117100	0565	1
6	135040	117100	0752	1
7	135040	117100	0980	1
8	135040	117100	1121	13
9	135040	117100	2681	13
10	135040	117100	2713	3825

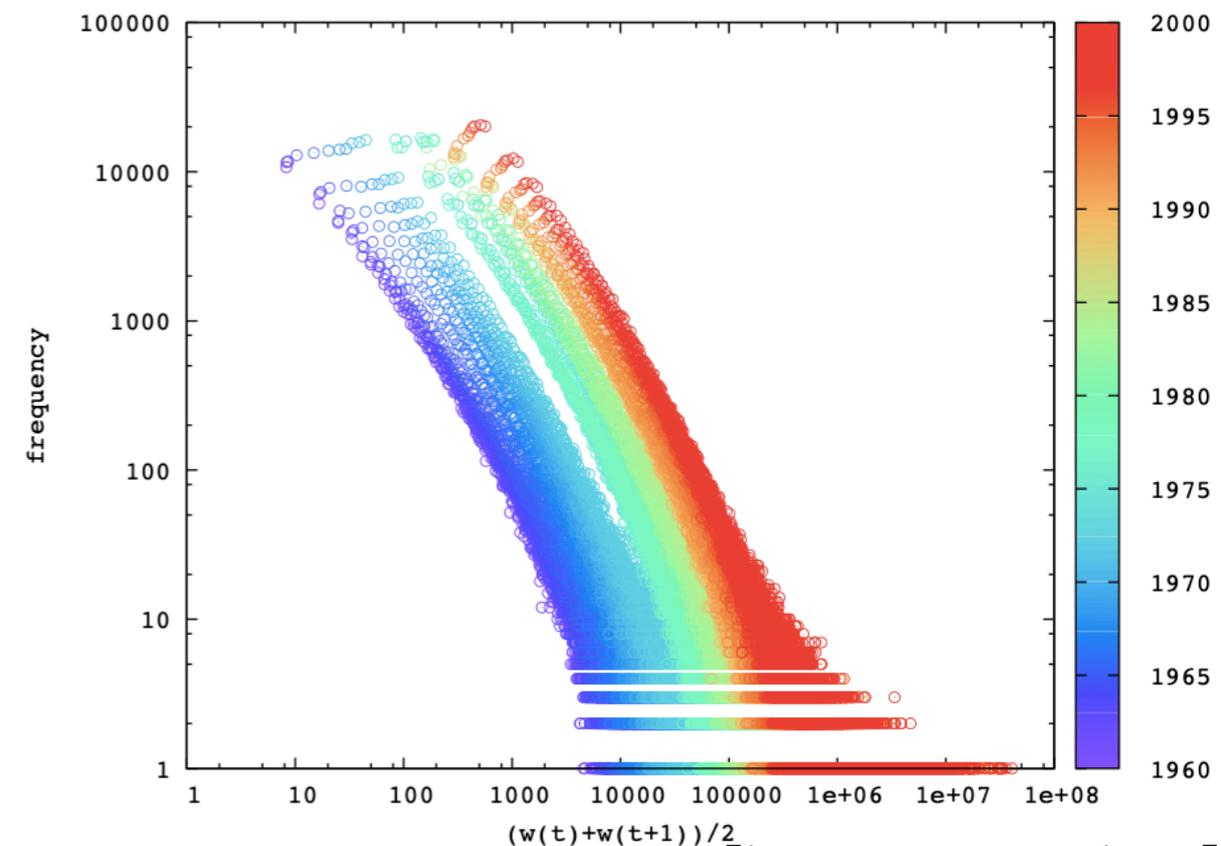
$$S(c_1) = \sum_{i=1}^{c_1(N)} w_i$$



$$w_N^i(t) = \frac{w_{Ex}^i(t) - w_{Im}^i(t)}{w_{Ex}^i(t) + w_{Im}^i(t)}$$



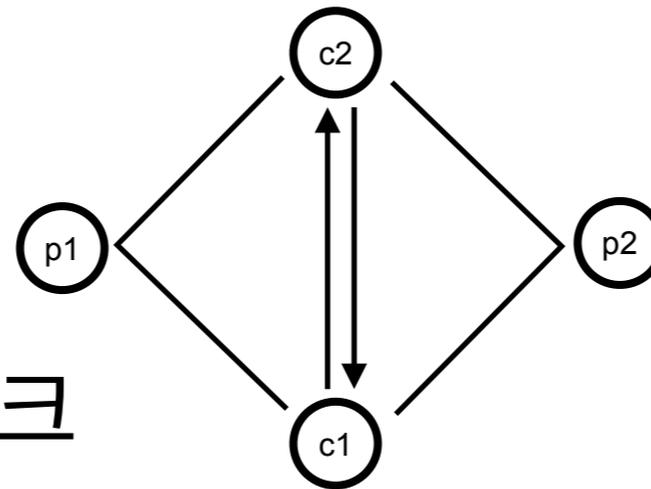
SITC Code	Product description
0	Food and live animals
1	Beverages and tobacco
2	Crude material, inedible, except fuels
3	Mineral fuels, lubricants and related materials
4	Animal and vegetable oils, fats and waxes
5	Chemicals and related products
6	Manufactured foods classified chiefly by material
7	Machinery and transport equipment
8	Miscellaneous manufactured articles
9	Commodities and transaction not classified elsewhere in the SITC



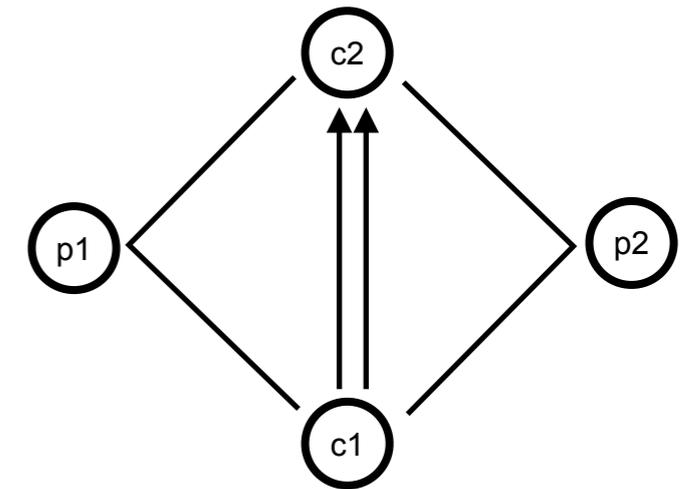
[in preparation]

Project: World trade network

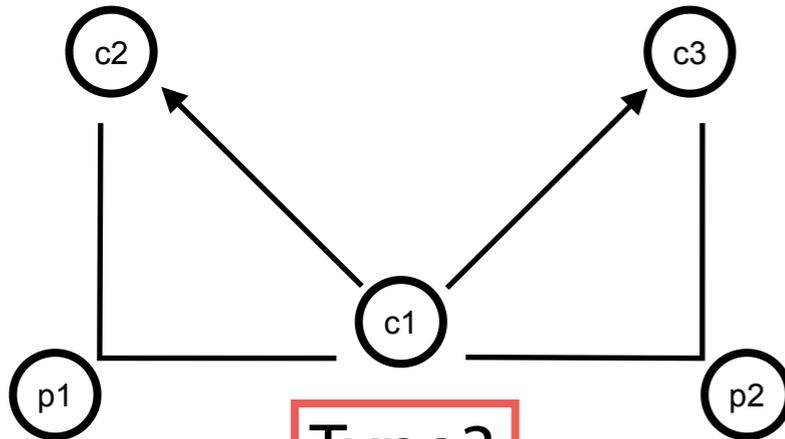
- Network Motif Analysis
- 6 types in 세계 무역 네트워크



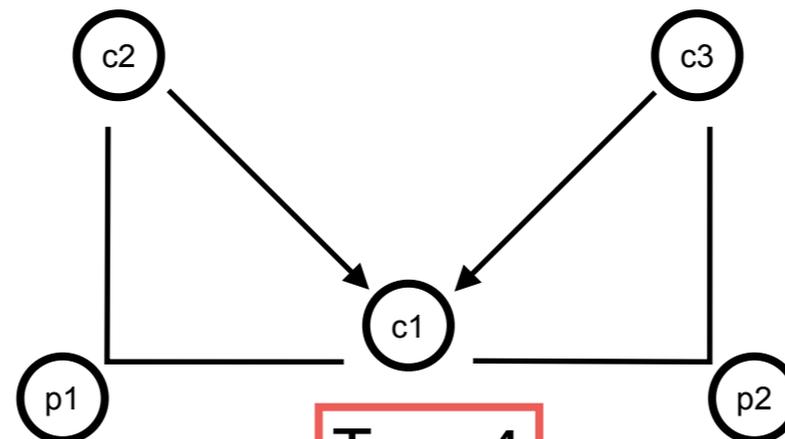
Type1



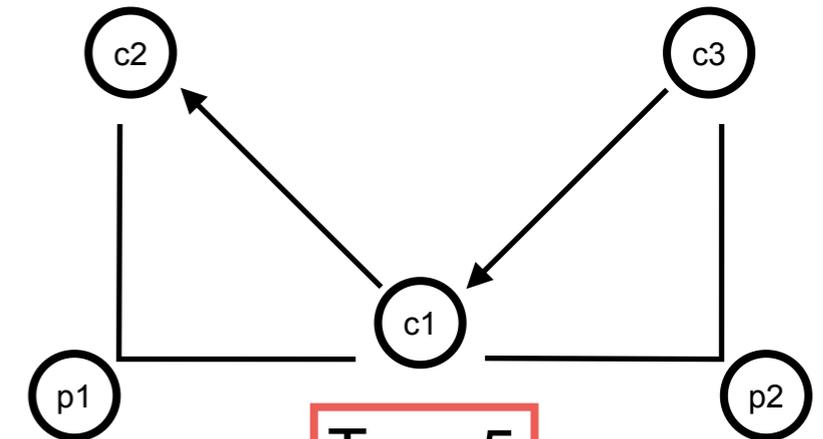
Type2



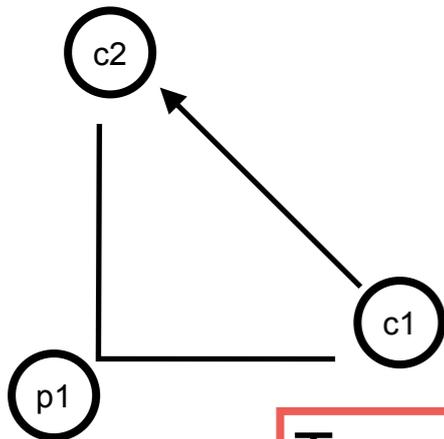
Type3



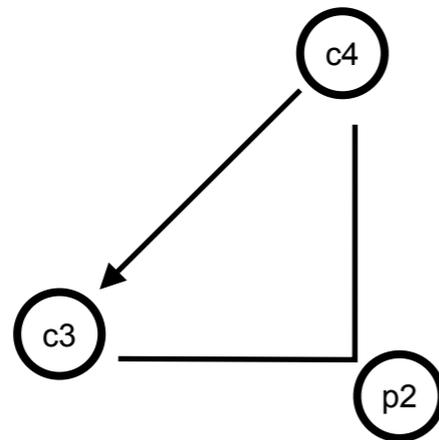
Type4



Type5



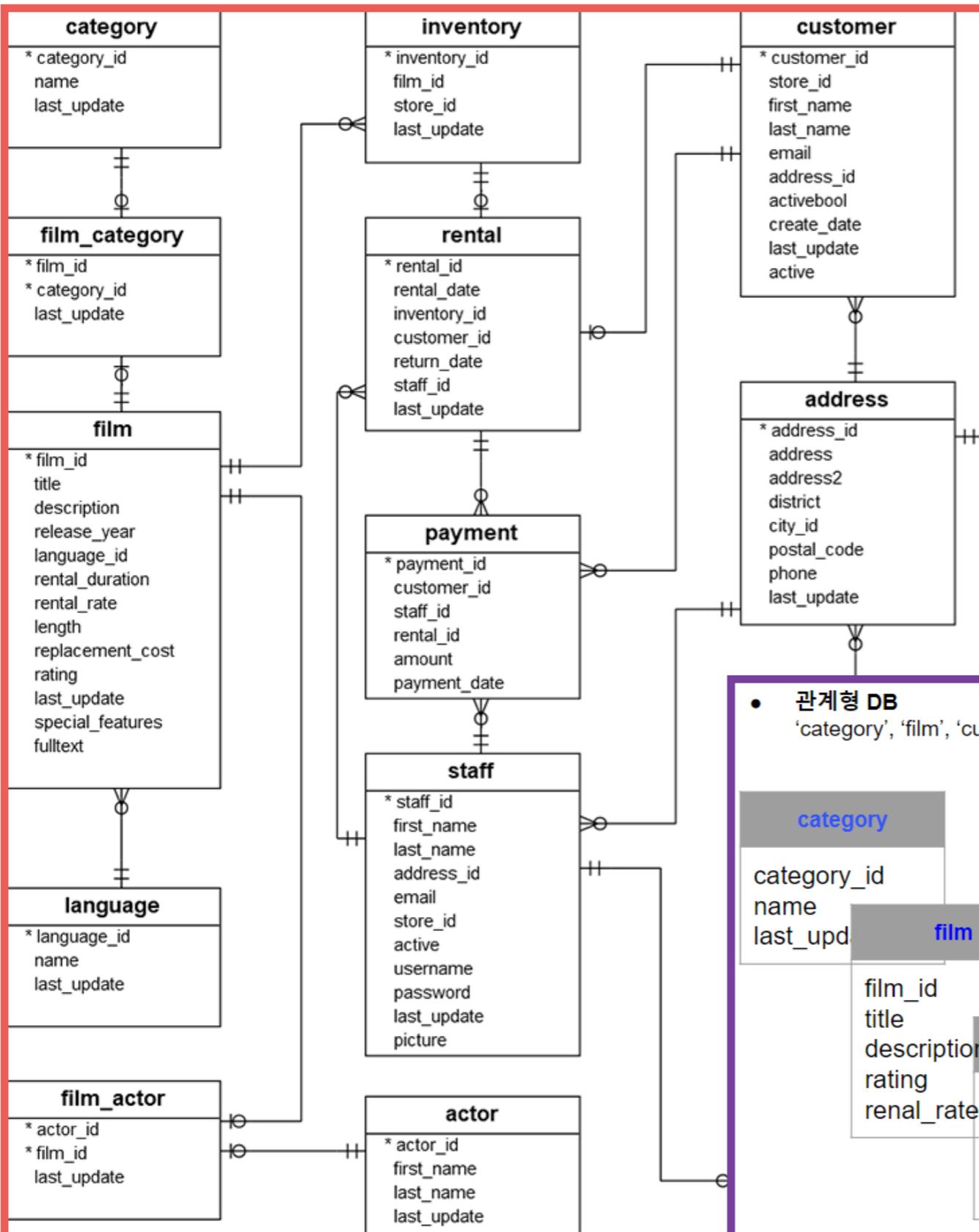
Type6



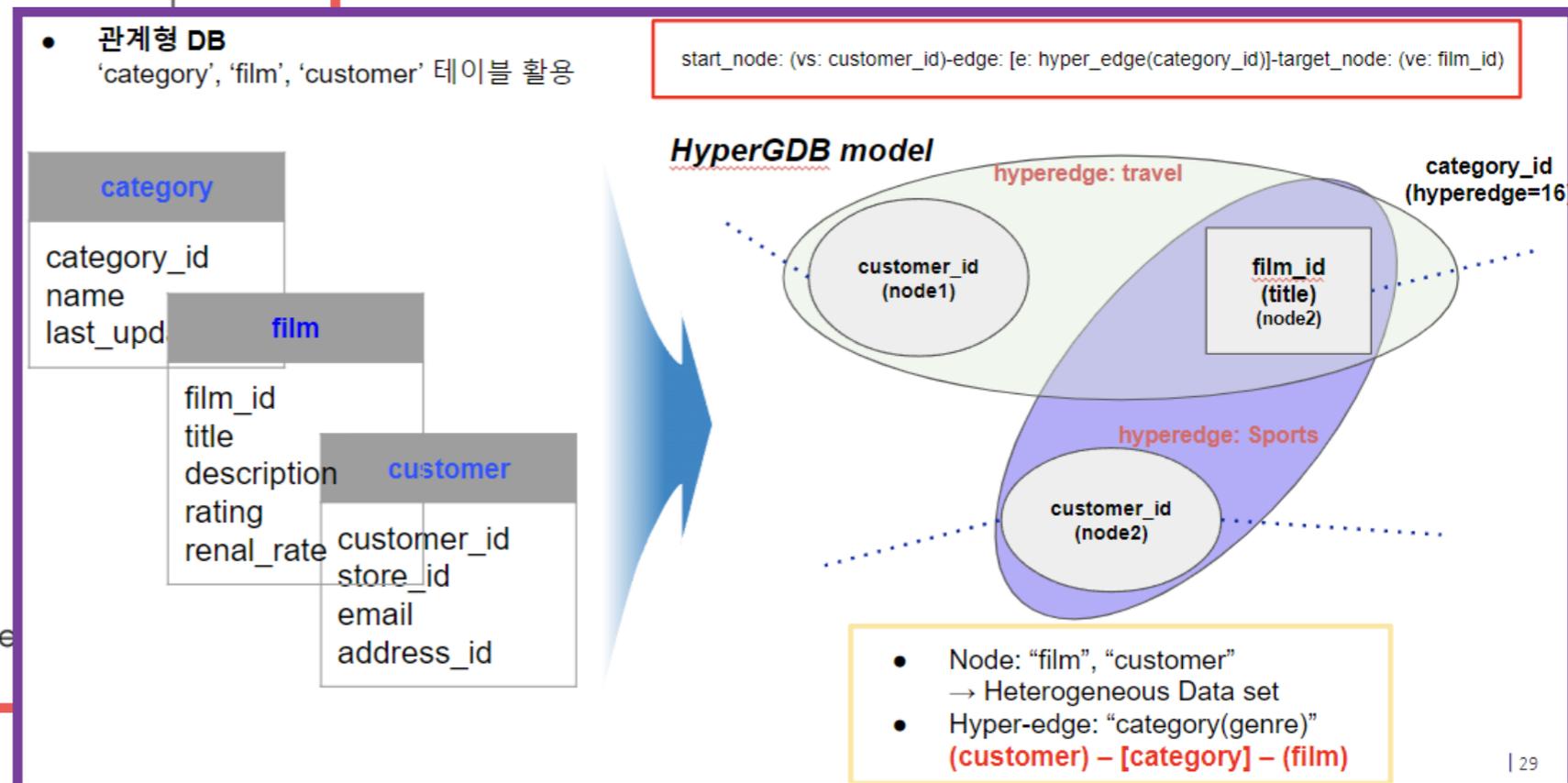
[in preparation]

Application to Hypergraphs: Database

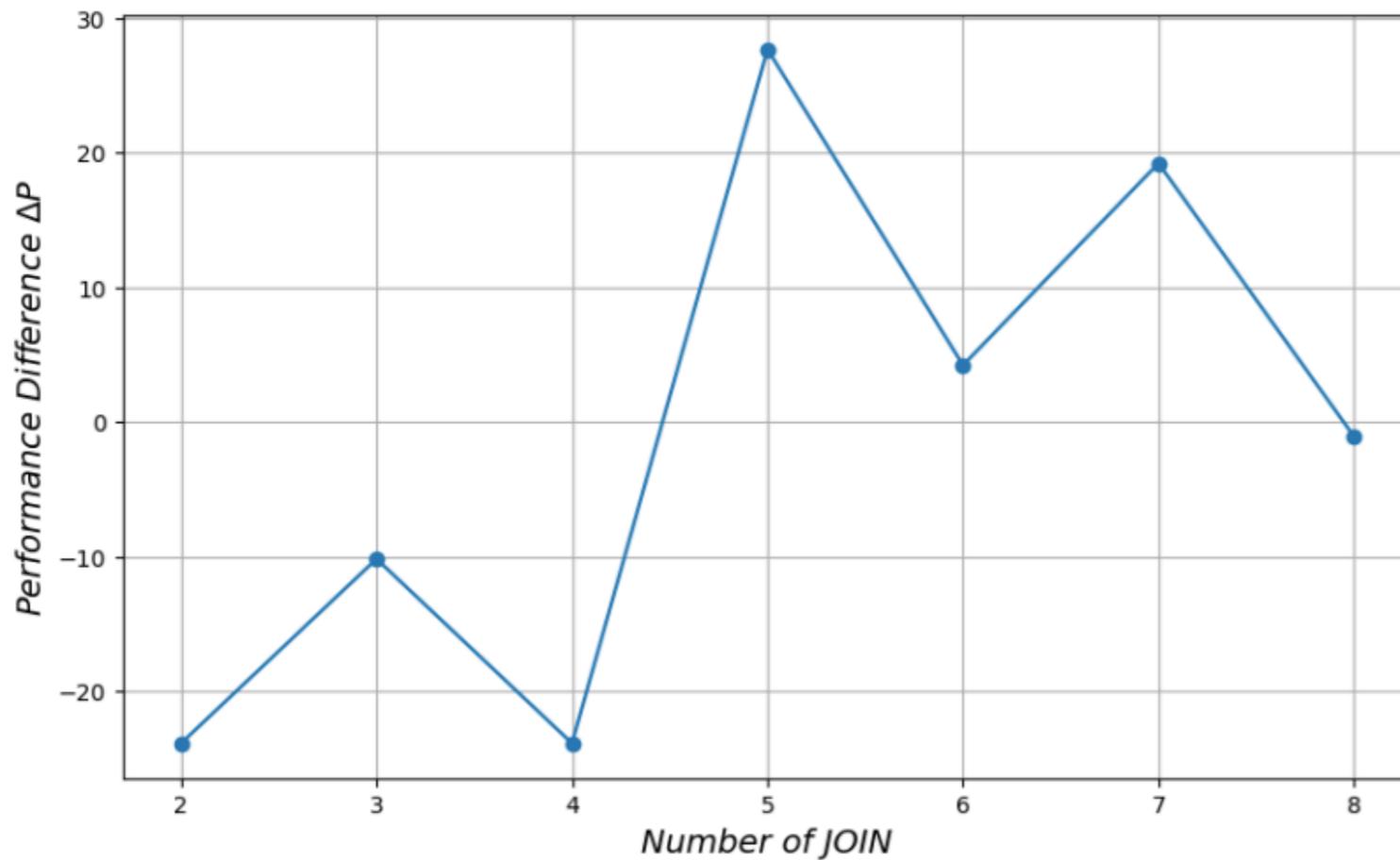
GDB to HGDB



- DVD 대여 서비스를 모델링한 데이터베이스 설계.
- 영화, 고객, 대여 등과 같은 테이블을 포함하여 관계를 나타냄.
- '카테고리(장르), '영화 제목', '고객' 테이블 활용.
- **노드: 고객, 대여한 dvd.**
- **하이퍼엣지: 대여한 dvd의 카테고리.**



HGDB Query Performance



*A Graph Extension for
PostgreSQL*



$$\Delta P = \frac{t_{RDB} - t_{HGDB}}{t_{RDB}} * 100$$

- HGDB는 엣지로 연결된 시스템에서 다양한 관계를 효과적으로 처리할 수 있는 데이터 구조를 제공.
- 하이퍼엣지와 연결성을 통해 다양한 관계를 다룰 수 있고 성능 이점을 얻을 수 있다.
- JOIN 연산이 적은 경우, 관계형 데이터베이스 (RDB)가 효과적일 수 있다.
- (하지만) JOIN 연산이 복잡하고 데이터의 연결 관계가 복잡한 경우, HGDB가 더 나은 성능을 보일 수 있다.

Summary I

특징	RDB	GDB	HGDB
정의	키(key)와 값(value)들의 관계를 테이블화 데이터베이스	노드와 엣지로 구성된 구조	노드와 하이퍼 엣지로 구성된 구조
노드	객체, 엔티티등 테이블 단위	개체, 지점, 사건등 다양한 요소	개체, 문서, 이미지등 공통 요소
관계 표현	여러 테이블의 JOIN 형태로 표현	노드 간의 관계를 엣지 로 표현	노드 간의 관계를 하이퍼 엣지 로 표현
방향성	단방향, 양방향 가능	단방향, 양방향 가능	단방향, 양방향 가능
속성	노드 및 테이블 에 속성 추가 기능	노드 및 엣지 에 속성 추가 기능	노드 및 하이퍼엣지 에 속성 추가 가능
예시	테이블의 1차원적 관계 표현	Social Net, KG(지식그래프)	웹 페이지, 협업 시스템, 웹 링크
장점	명확한 스키마로 구조 변화와 데이터 관리에 적합	복잡한 관계의 데이터 관리와 표현에 적합	복잡한 다중 관계에 대한 데이터 관리와 표현에 적합

Summary II

[데이터 분석 관점]

- **다양한 관계 네트워크의 모델링:** (하이퍼 그래프 데이터) 복잡한 관계 정보에 대한 네트워크를 모델링하고 분석.
- **중요한 중심성 노드 식별:** 중심성을 정량적으로 계산하여 중심성에 따른 노드를 구분. 중심성 높은 노드는 네트워크 내에서 핵심 역할을 하거나 정보 전파나 영향력 전파에 중요한 역할.
- **커뮤니티 탐지:** 서로 관련된 노드의 그룹, 즉 커뮤니티를 탐지하여 유사한 관심사나 행동을 가진 그룹을 식별하고 이해.
- **경로 분석:** 노드 간의 다양한 경로를 분석. 이를 통해 정보의 전파 경로나 상호 작용을 이해하고 최적 경로를 찾는다.
- **시간에 따른 변화 분석:** 시간에 따라 네트워크 구조나 관계가 어떻게 변화하는지 이해한다. 이를 통해 동적인 관계의 흐름을 파악하고 예측.

[비즈니스 분석 관점]

- **이상 탐지(FDS):** 네트워크 내의 이상한 패턴을 탐지. 이상한 행동이나 연결을 감지하여 보안 및 사기 탐지 등 다양한 응용 분야에 활용. (Find Cyclic Loop)
- **개인화 추천:** 초개인화된 추천 시스템을 구축. 사용자의 관심사와 유사한 다른 사용자나 리소스를 찾아 추천.
- **의사 결정 지원:** 네트워크 관계를 통해 전략을 개발하고 리더십을 제시하는 데 도움.

[시각화 관점]

- **시각화를 통한 인사이트 제공:** 데이터를 시각화하여 하이퍼그래프의 복잡한 구조를 이해하기 쉽게 만들어 인사이트를 제공하고 결과를 효과적으로 전달.

[Sang-Hwan Gwak 곽상환]

PhD in Physics (Statistical Physics) under prof "Kwang-II Goh(고광일)" from Korea Univ.

[Contact address]

E-mail: gwak.sanghwan(at)gmail.com

[Research interests]

Graph Database, Network Science and theory, Complex systems
-> Structural and dynamical properties of complex networks.

GUG 2nd Seminar에 참석해 주셔서 감사합니다.

그리고 '정태'님 감사합니다.

Any questions are welcome!!

(gwak.sanghwan@gmail.com)