

COURSE : FUNDAMENTALS OF COMPUTER SCIENCE

LECTURE 10: “NETWORK ANALYSIS, THEORY AND PERFORMANCE”

Instructor:
PhD, Associate Professor
Tuyatsetseg Badarch

Сүлжээнүүдийн жишээ



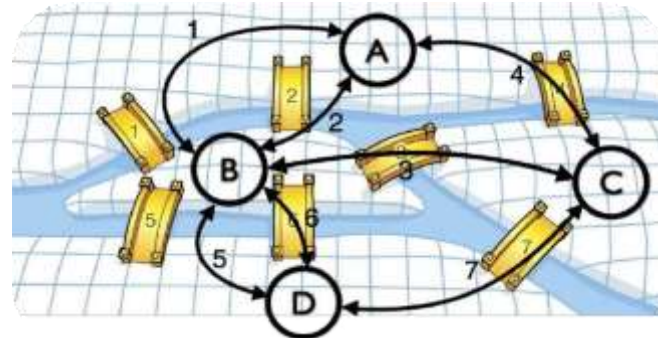
Кёнигсбергийн гүүр

"Өөрийн оюун ухааныг ашиглах эр зоригтой бай."

Графын үндсэн ойлголт, хэмжүүр, алгоритмууд

Networkx комплекс буюу томоохон хэмжээний ойлгоход төвөгтэй сүлжээнүүд дээр анализ, шинжилгээ хийхэд зориулагдсан пайтон хэл дээр бичигдсэн сан юм.

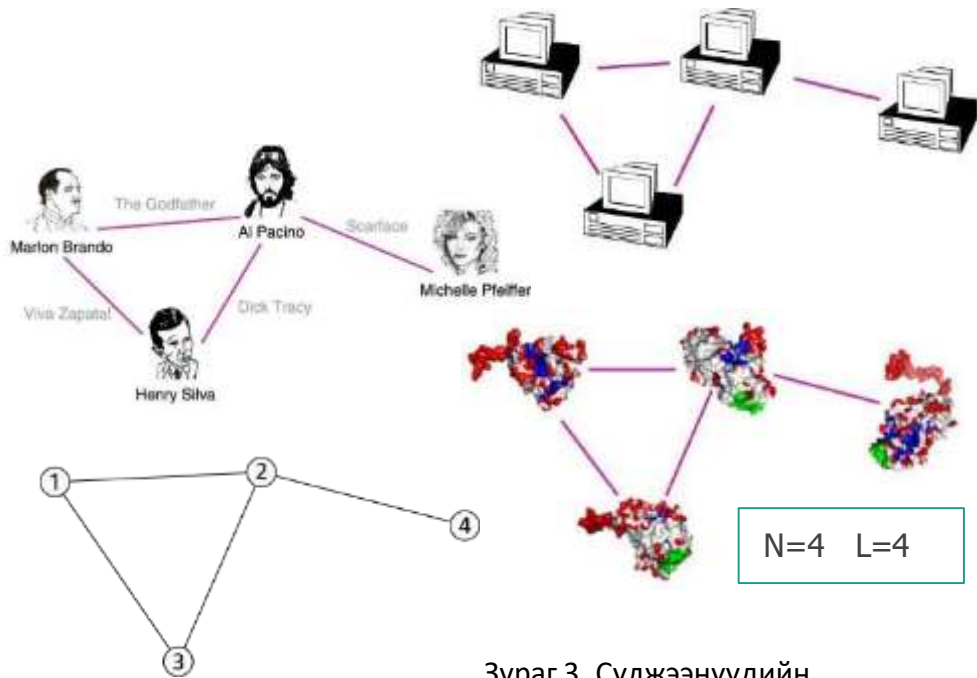
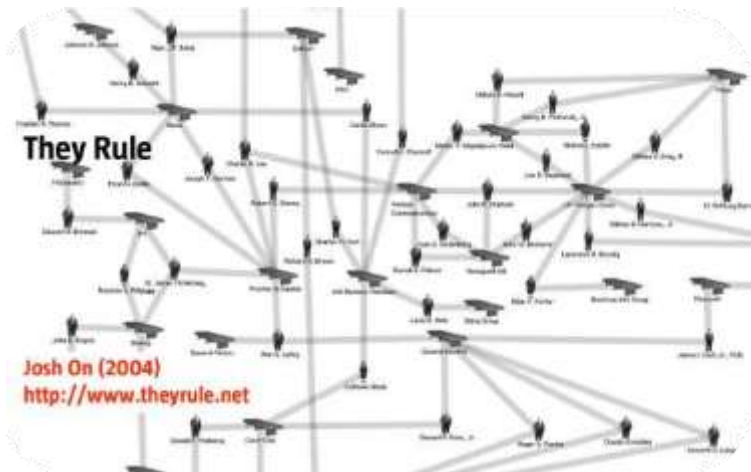
Энэхүү лабораторын ажлаараа дээрх сангийн тухай болон үүн дээр хийгддэг үндсэн гол функцүүдийн талаар судална. Хэрэгцээт сангуудаа суулгаж эхлэхээс эхлэнэ. Жишээ нь дараах зурагт өгөгдсөн байдлаар.



Зураг 2. Сүлжээнүүдийн жишээ, Кенигсбергийн гүүр

Сүлжээнүүд ба системүүд

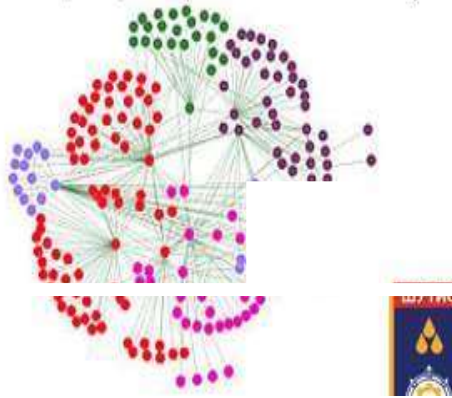
Social - Нийгмийн, олон нийтийн, сүргийн
Network - Сүлжээ, тор
Social network - Нийгмийн сүлжээ, олон нийтийн сүлжээ



Зураг 3. Сүлжээнүүдийн жишээгээр үзүүлэх

Нийгмийн сүлжээний тодорхойлолт

- Нийгмийн сүлжээ гэдэг нь хосолсон холбоос бүхий (dyadic нийгмийн төлөөлөгч (social actor) нараас бүтсэн төлөөлөгч бүр нь өөр хоорондоо холбогдсон нийгмийн бүтэцийг (social structure) хэлнэ.
- dyadic ties - хосолсон холбоос
- social actor - нийгмийн төлөөлөгч
- social structure - нийгмийн бүтэц



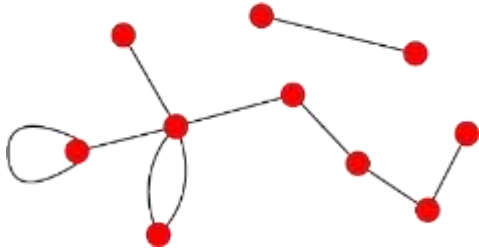
Зураг 4. Нийгмийн сүлжээний тодорхойлолт

Чиглэл

Чиглэлгүй граф

Undirected graphs

Links: undirected (*symmetrical*)



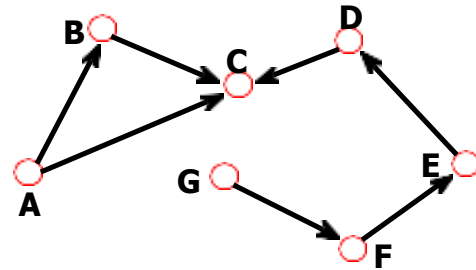
Examples of Undirected links

Co-authorship links
Actor network
Protein interactions

Чиглэлт граф

Directed graphs (DiGraphs)

Links: directed (*arcs*).



Example of Directed links

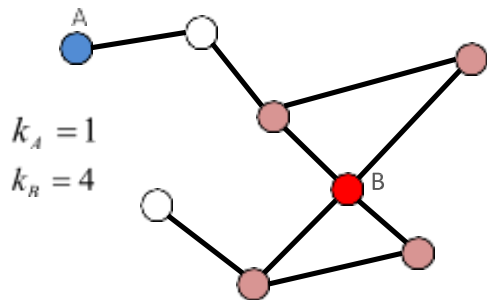
URLs on the www
Phone calls
Metabolic reactions

Зураг 5. Сүлжээнүүдийг
графын онолоор загварчлах

NETWORK	NODES	LINKS	DIRECTED		N	L
			UNDIRECTED			
Internet	Routers	Internet connections	Undirected		192,244	609,066
WWW	Webpages	Links	Directed		325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected		4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed		36,595	91,826
Email	Email addresses	Emails	Directed		57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected		23,133	93,439
Actor Network	Actors	Co-acting	Undirected		702,388	29,397,908
Citation Network	Paper	Citations	Directed		449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed		1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected		2,018	2,930

Зэрэг, Дундаж зэрэг, Зэргийн тархалт

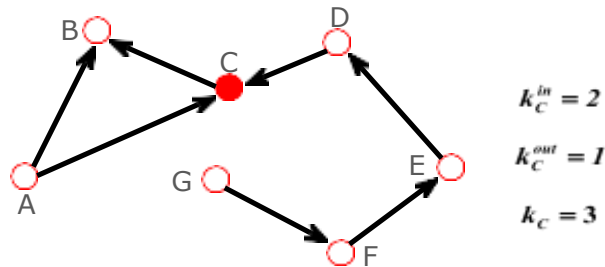
Чиглэлгүй



Чиглэлт (DiGraphs)

in-degree

out-degree.



Зураг 6. Сүлжээнүүдийн загварчлал

Загварчлал

Дөрвөн параметр : N утга x_1, \dots, x_n

Дундаж утга

$$\langle x \rangle = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{1}{N} \sum_{i=1}^n x_i$$

n^{th} момент:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \dots + x_n^n}{N} = \frac{1}{N} \sum_{k=1}^N x_k^n$$

Стандарт хазайлт :

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2}$$

X -ын тархалт

$$p_x = \frac{1}{N} \sum_i \delta_{x, x_i}$$

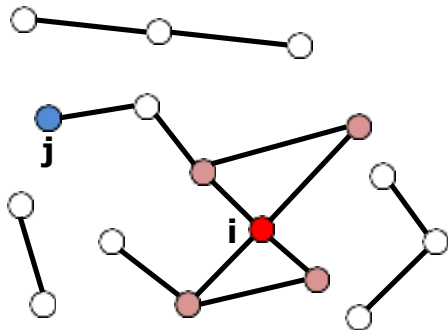
P_x :

$$\sum_x p_x = 1 \left(\int p_x dx = 1 \right)$$

Дундаж зэрэг

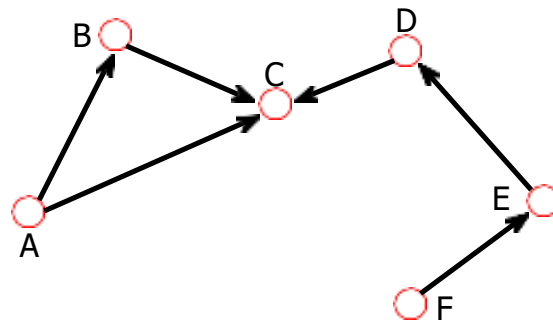
Зураг 7. Сүлжээнүүдийн загварчлал

Чиглэлгүй



$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i \quad \langle k \rangle \equiv \frac{2L}{N}$$

Чиглэлт (DiGraphs)



$$\langle k^{in} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{out},$$

$$\langle k^{in} \rangle = \langle k^{out} \rangle \quad \langle k \rangle = \frac{L}{N}$$

Практик жишээнүүд

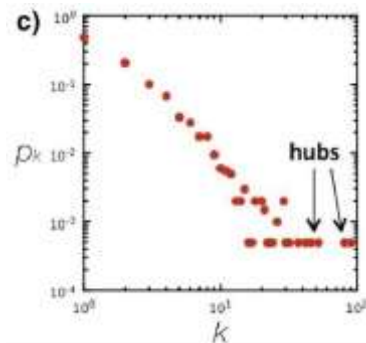
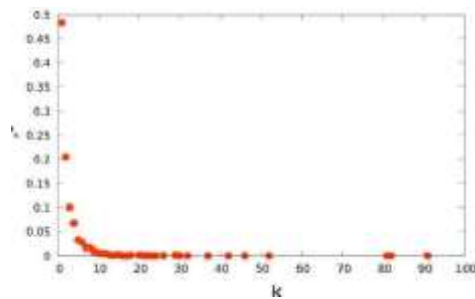
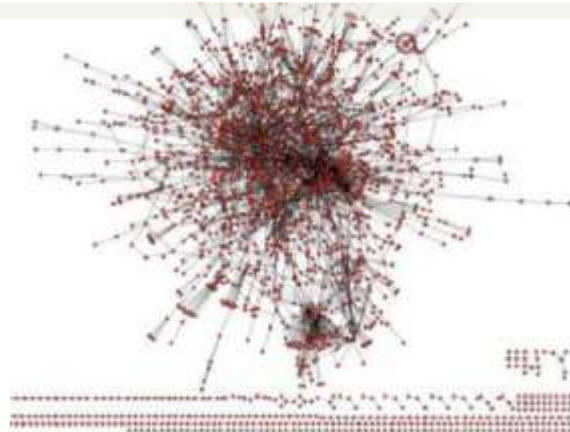
NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	$\langle k \rangle$
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Зэргийн тархалт

$P(k)$: магадлал

$N_k = k$ зэрэгтэй оройн тоо

$P(k) = N_k / N \rightarrow$ плот



Зураг 8. Сүлжээнүүдийн зэргийн ТАРХАЛТ

Зэргийн тархалт

Тоон : p_k - k зэрэгтэй оройн байх магадлал .

Аналог: $p(k)$ is дээрх утгын pdf

$$\int_{k_1}^{k_2} p(k) dk$$

Магадлал k_1 ба k_2 .

Нормчлолын нөхцөл :

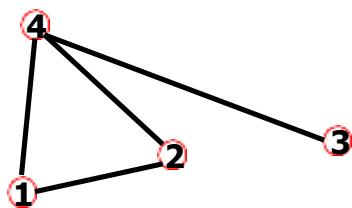
$$\sum_a^{\infty} p_k = 1 \quad \int_{k_{min}} p(k) dk = 1$$

k_{min} хамгийн бага зэрэг

Хөршийн матриц

Зураг 9. Хөршийн матриц

Чиглэлгүй



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

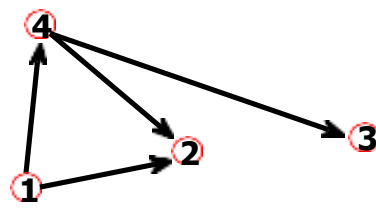
$$k_i = \sum_{j=1}^N A_{ij}$$

$$k_j = \sum_{i=1}^N A_{ij}$$

$$A_{ij} = A_{ji} \\ A_{ii} = 0$$

$$L = \frac{1}{2} \sum_{i=1}^N k_i = \frac{1}{2} \sum_{ij} A_{ij}$$

Чиглэлтэй (DiGraphs)



$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$k_i^{in} = \sum_{j=1}^N A_{ij}$$

$$k_j^{out} = \sum_{i=1}^N A_{ij}$$

$$A_{ij} \neq A_{ji} \\ A_{ii} = 0$$

$$L = \sum_{i=1}^N k_i^{in} = \sum_{j=1}^N k_j^{out} = \sum_{i,j} A_{ij}$$

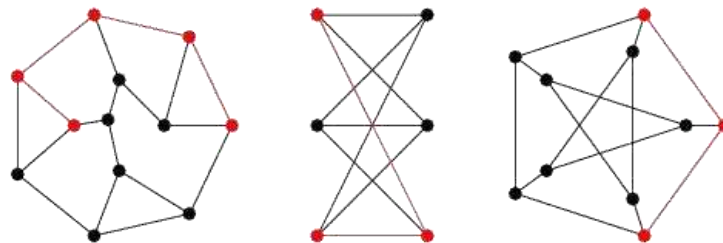
Зам болон холбоосууд

Замууд

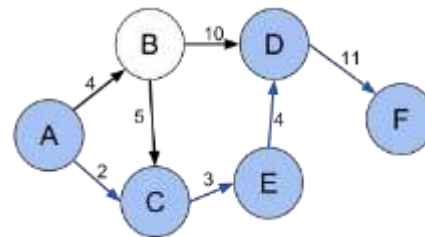
$$P_n = \{i_0, i_1, i_2, \dots, i_n\}$$

$$P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_{n-1}, i_n)\}$$

Зураг 10. Сүлжээнүүдийн
чиглэлгүй графын зам



Чиглэлгүй графын зам .



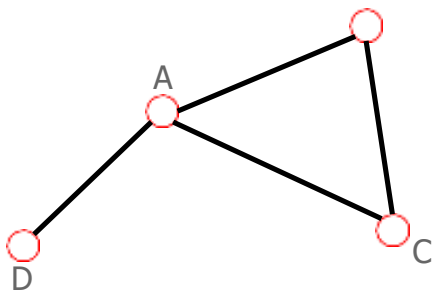
Чиглэлт графын зам

Граф графын үзүүлэлтүүд

Граф дахь зай

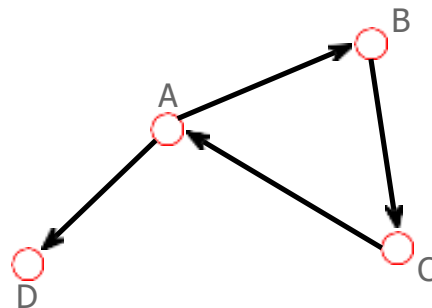
Чиглэлгүй граф

The *distance* (хамгийн богино зам, жеодик зам)



Холбогдоогүй зангилаа буюу оройнуудын хоорондох зай тодорхой бус

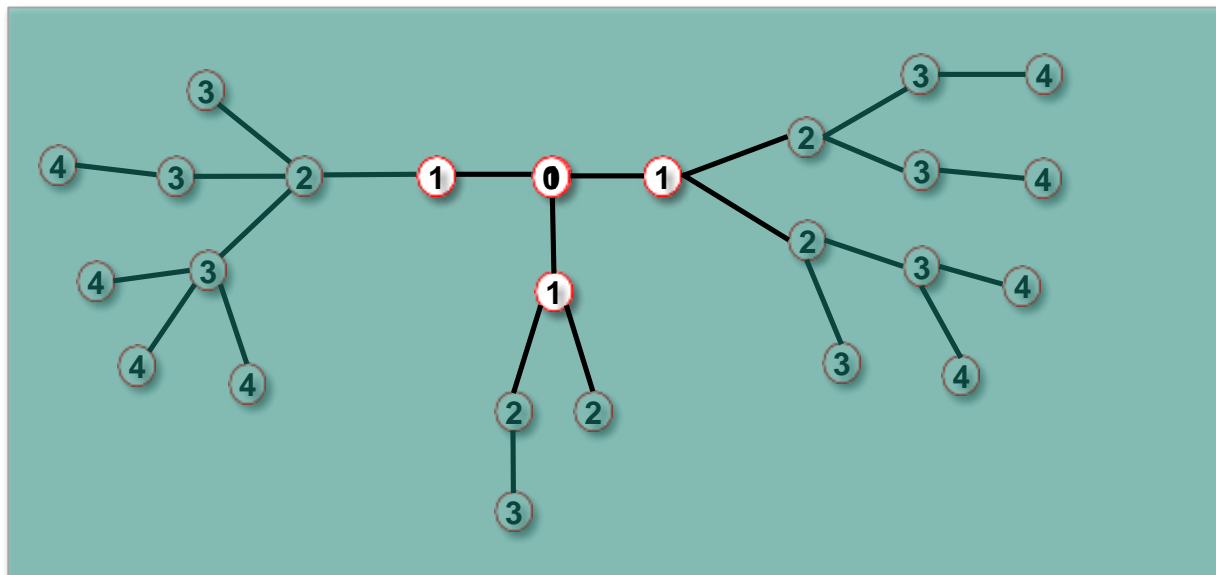
Чиглэлт (DiGraphs)



Зураг 11. Сүлжээнүүдийн графын зам

Графын онолын алгоритмын жишээ

Зай олох: BFS

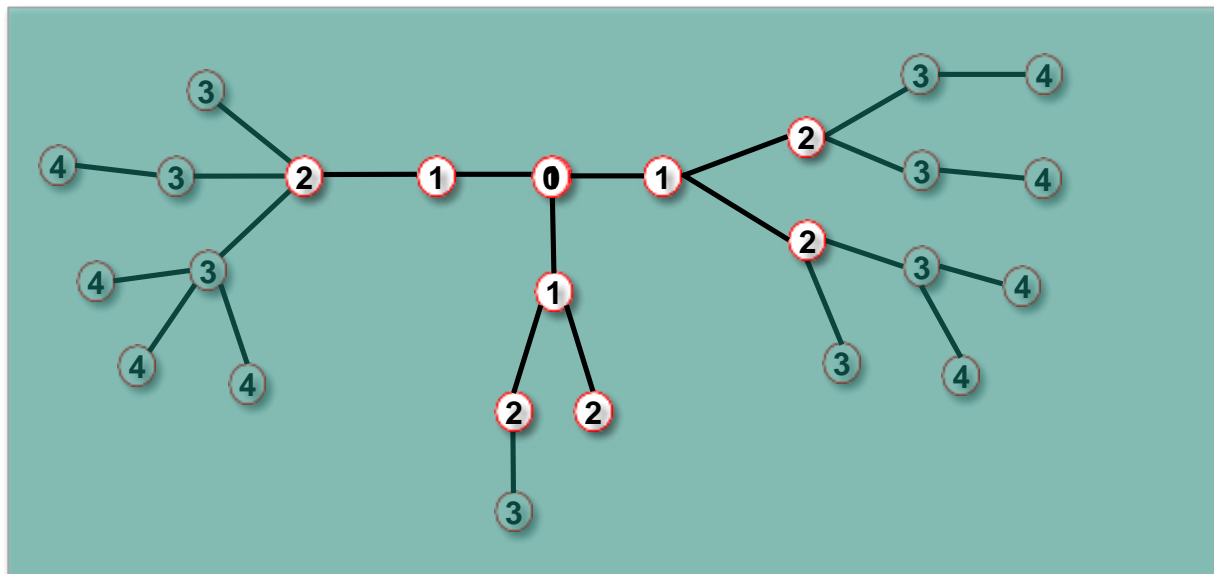


2. Дараагийн удаа 1 гэсэн түвшинтэй оройнууд байх бөгөөд эдгээр нь өмнөх 0 түвшиний оройтой шууд ирмэгээр холбогдсон буюу хөрш оройнууд байх ба бүгдийг 1ээр тэмдэглэж дараалалд оруулна.

Зураг 13. Сүлжээнүүдийн орой буюу зангилаануудын зайг олох

Графын онолын алгоритмын жишээ

Зай олох : BFS

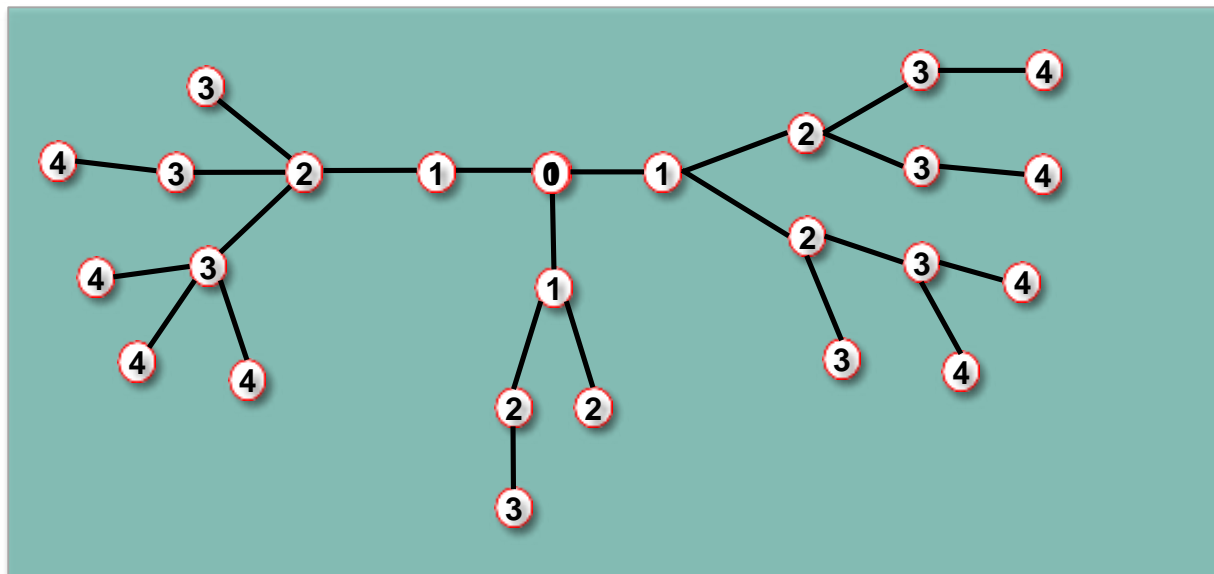


3. Өмнөх алхамд үүсгэсэн дарааллаас эхний элемент буюу оройг авч түүнтэй холбогдсон дараагийн тэмдэглээгүй оройнуудыг олж 2 гэж тэмдэглээд тэдгээрийг мөн тухайн дараалалдаа нэмнэ.

Зураг 14. Сүлжээнүүдийн орой буюу зангилаануудын зайг олох

Графын онолын алгоритмын жишээ

Зай олох : BFS



Дараалалд дахиж сонгогдох элемент байхгүй болтол уг үйлдлээ давтана. Тэгэхээр 0 болон 4 гэж тэмдэглэсэн

оройнуудын хоорондох зай нь тухайн

өгөгдсөн шошго буюу 4тэй тэнцүү. Хэрэв ямарч шошго тавигдаагүй бол уг хоёр орой холбоогүй. ооронд зай нь хязгааргүй

Зураг 15. Сүлжээнүүдийн орой буюу зангилаануудын зайг олох

Графын онолын алгоритмын жишээ

Диаметр болон дундаж зай

Диаметр (d_{\max}):

Хамгийн хол зай

Графын диаметр гэдэг нь тухайн граф дахь аль нэг хоёр оройн хоорондох хамгийн урт зайтай тэнцүү.

Дундаж зай:

$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} d_{ij}$$

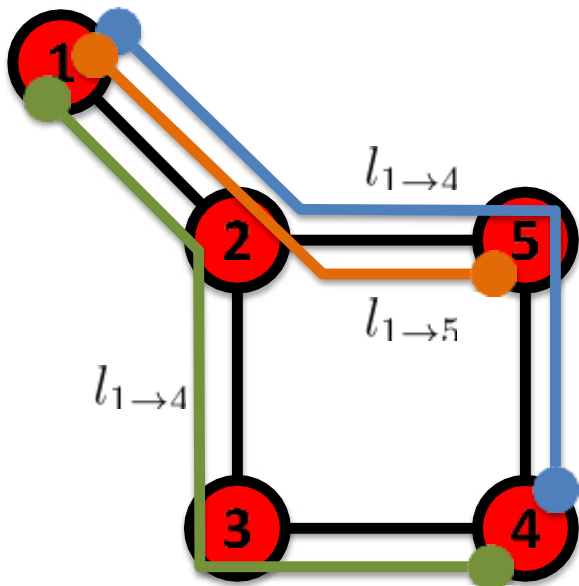
Чиглэлгүй: $d_{ij} = d_{ji}$

$$\langle d \rangle \equiv \frac{1}{L_{\max}} \sum_{i,j > i} d_{ij}$$

Хамгийн бага зай

МИН богино зай

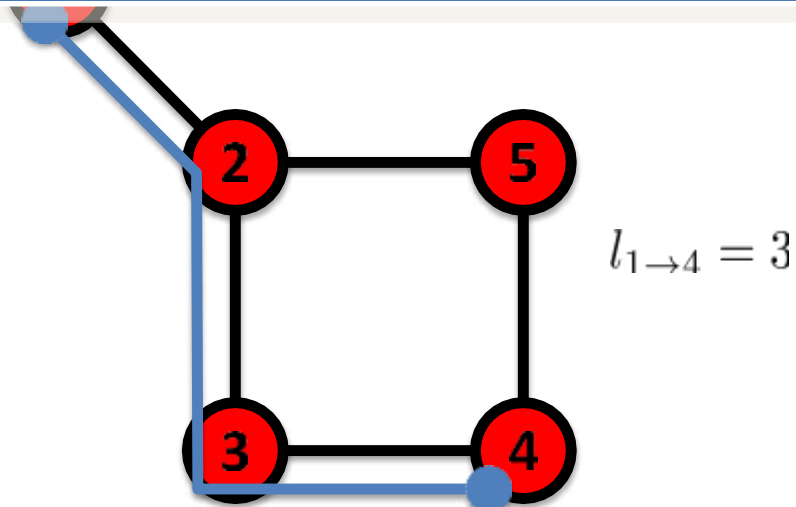
Хоёр оройн хоорондох хамгийн богино урттай замыг тухайн оройнуудын хоорондох богино зам гэнэ.



$$l_{1 \rightarrow 4} = 3$$

$$l_{1 \rightarrow 5} = 2$$

Диаметр бол хамгийн тухайн граф дахь хамгийн урт зам. Энэхүү жишээн дээр 1 болон 4 гэсэн оройнуудыг холбосон хамгийн богино зам болж байна.



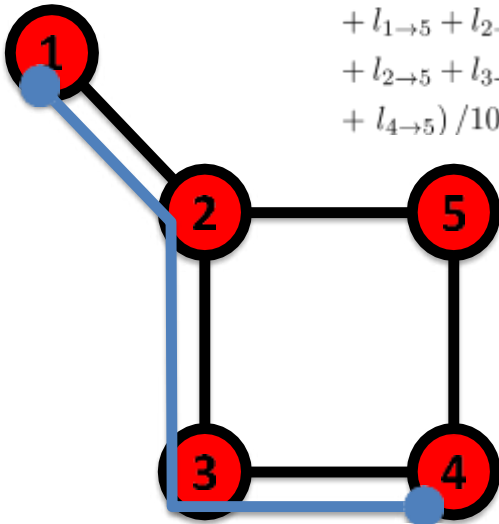
Зураг 16. Сүлжээнүүдийн орой буюу зангилаануудын хамгийн богино зайг олох

Дундаж замын урт

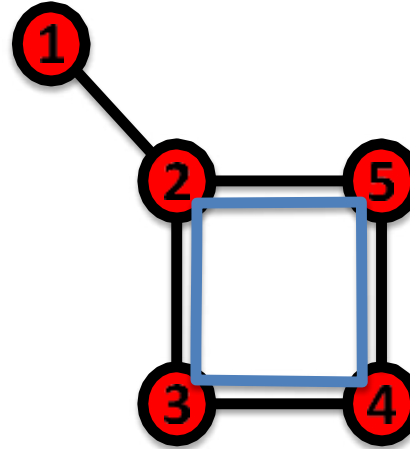
Дундаж замын урт

Граф дахь бүх хос оройг холбосон богино замуудыг дундажласан утгыг дундаж богино зам гэнэ.

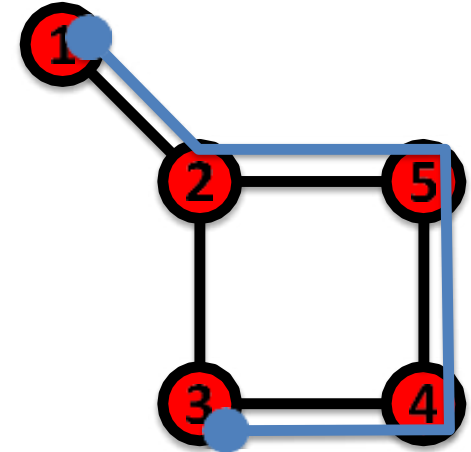
$$\begin{aligned} & (l_{1 \rightarrow 2} + l_{1 \rightarrow 3} + l_{1 \rightarrow 4} + \\ & + l_{1 \rightarrow 5} + l_{2 \rightarrow 3} + l_{2 \rightarrow 4} + \\ & + l_{2 \rightarrow 5} + l_{3 \rightarrow 4} + l_{3 \rightarrow 5} + \\ & + l_{4 \rightarrow 5}) / 10 = 1.6 \end{aligned}$$



Эхлэл болон төгсгөлийн оройн ижил замыг цикл гэнэ.

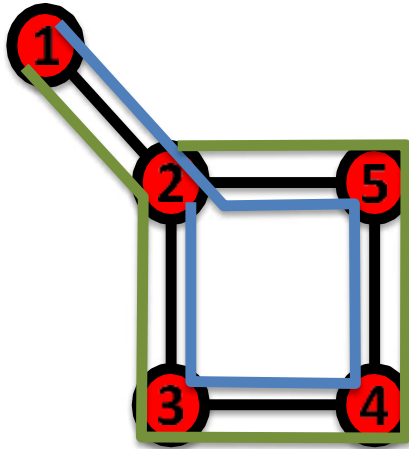


Огтлолцол үүсэхгүй замыг өөрөөс зугтах зам гэнэ.



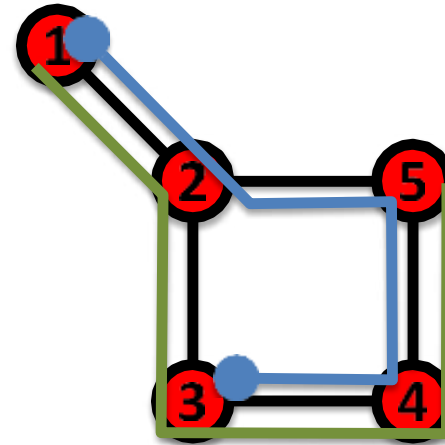
Зураг 17. Сүлжээнүүдийн орой буюу зангилаануудын хамгийн богино зайг олох

Дундаж замын урт



Eulerian Path/Cycle

Бүх ирмэгүүдээр нэг удаа зорчидог замыг Эйлерийн зам.



Hamiltonian Path/Cycle

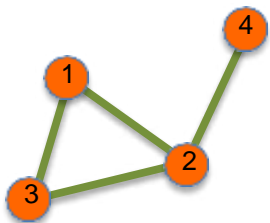
Бүх оройгоор яг нэг удаа зочилдог замыг Хамилтон зам/цикл.

Зураг 18. Сүлжээнүүдийн орой буюу зангилаануудын хамгийн богино зайг олох

Сүлжээний чиглэл

Чиглэл

Чиглэлгүй

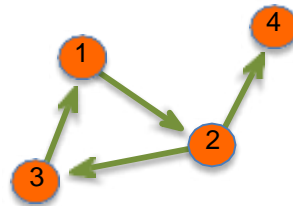


$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \\ L = \frac{1}{2} \sum_{i,j=1}^N A_{ij}$$

$$A_{ij} = A_{ji} \\ \langle k \rangle = \frac{2L}{N}$$

Чиглэлтэй



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \\ L = \sum_{i,j=1}^N A_{ij}$$

$$A_{ij} \neq A_{ji} \\ \langle k \rangle = \frac{L}{N}$$

Зураг 19. Сүлжээнүүдийн орой буюу зангилаануудын хамгийн богино зайг олох

Хэмжээ хамаарахгүй сүлжээ

Оройн зэргийн тархалт нь power-law тархалтаар тодорхойлогдох сүлжээг Scale-Free сүлжээ гэнэ.

$$P(k) \sim Ck^{-\gamma} = C \frac{1}{k^{\gamma}}$$

Бүх оройн зэрэг эерэг бүхэл тоо байх бөгөөд эндээс дискрет нормальчлал хийснээр тухайн оройг яг k холбоос үүсгэсэн байх магадлалыг олж болно.

$$p_k = Ck^{-\gamma}, \quad \sum_{k=1}^{\infty} p_k = 1$$
$$C \sum_{k=1}^{\infty} k^{-\gamma} = 1$$
$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}, \quad p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

P_k

Аналог

зэргүүд нь ямар нэг эерэг бодит утга авдаг тохиолдолд

$$p(k) = Ck^{-\gamma} \quad \int_{k_{\min}}^{\infty} p(k)dk = 1$$
$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$
$$p(k) = (\gamma - 1)k_{\min}^{\gamma-1} k^{-\gamma}$$

$$\int_{k_1}^{k_2} p(k)dk$$

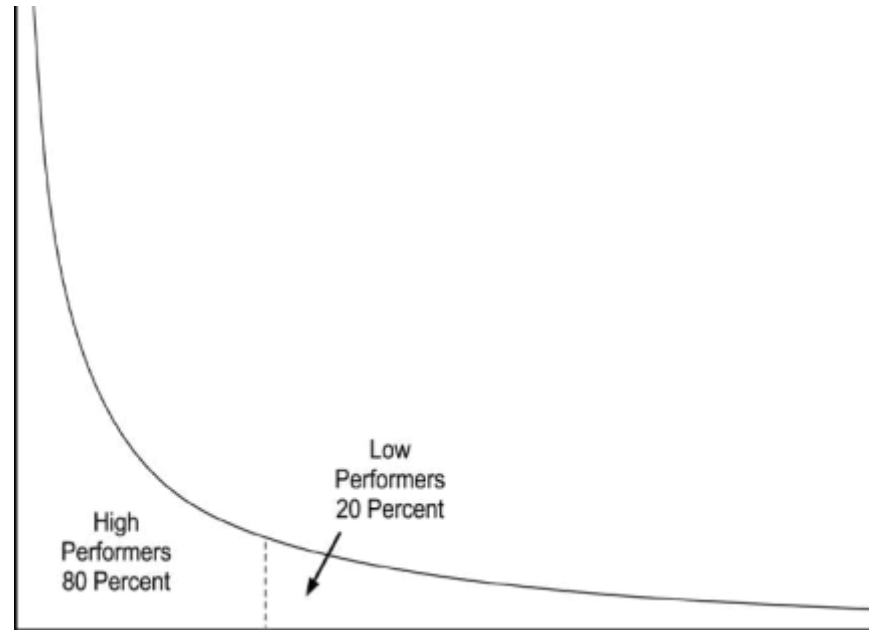
Парето хууль

Паретогийн хууль буюу 80/20 дүрэм

80/20 Rule

Vilfredo Federico Damaso Pareto (1848–1923)
Pareto distribution

Зураг 20. Парето хууль



Сүлжээний зангилаанууд

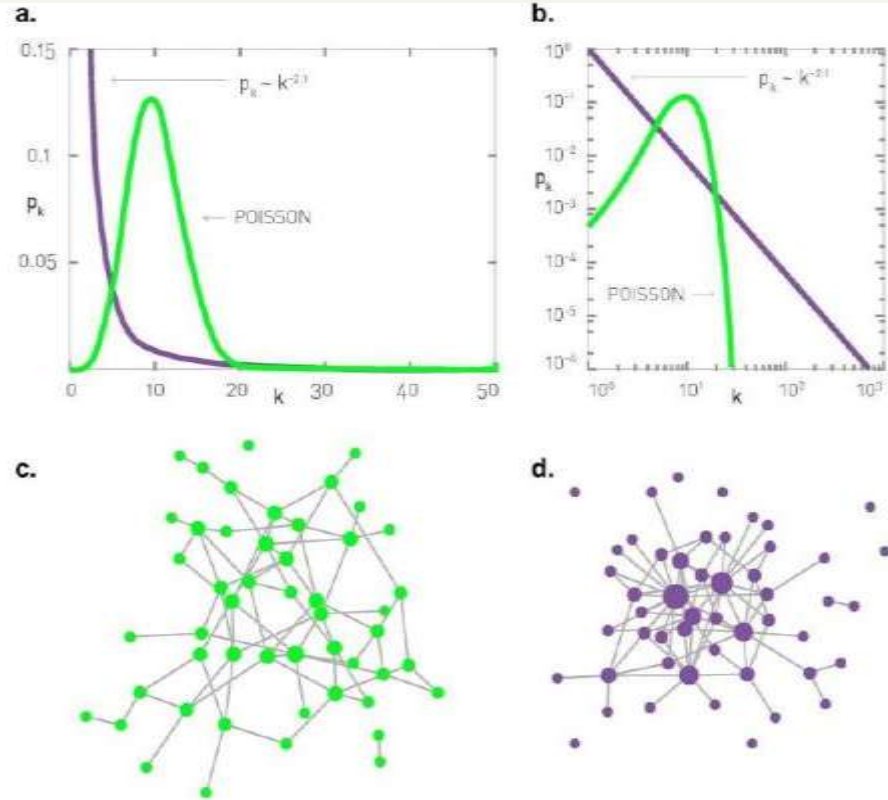
Зангилаанууд

Санамсаргүй ба хэмжээ чөлөөтэй сүлжээнүүдийн хоорондо ялгаа
ЗЭРГИЙН ТАРХАЛТЫН СҮҮЛЭЭР ЯЛГААТАЙ. ЭНЭ НЬ МАГАДЛАЛЫН k -ИХ ЗЭРГИЙГ ТОДОРХОЙЛДОГ.

МАШ ОЛОН ТООНЫ БАГА ЗЭРГИЙН ОРОЙ БУЮУ ЗАНГИЛААТАЙ, ИХЭНХ
ЗАНГИЛААНУУД НЬ САНАМСАРГҮЙ СҮЛЖЭЭНД ОЛДДОГҮЙ.

К ЗЭРЭГТЭЙ ОРОЙ БУЮУ ЗАНГИЛААНУУДЫН ИЛҮҮДЭЛТЭЙ СҮЛЖЭЭ

ИХ ЗЭРЭГТЭЙ ОРОЙ БУЮУ ЗАНГИЛАА НЬ ОЛОН БАЙДАГ СҮЛЖЭЭ.



Зураг 21. Сүлжээнүүдийн орой буюу зангилаанууд

Сүлжээнүүдийн харьцуулалт

Сүлжээнүүдийн харьцуулалт



Зангилааны тоо	$N = t$
Шугамын тоо	$N = mt$
дундаж зэрэг	$\langle k \rangle = 2m$
Зэргийн тархалт	$P(k) \sim Ck^{-\gamma}$
Кластерлэлт	$\frac{m}{4} \frac{(\ln N)^2}{N}$
Замын урт	$\frac{\ln N}{\ln \ln N}$

Network	Зэргийн тархалт	Замын урт	Кластерын сүлжээ
Бодит сүлжээ	Өргөн зурвас	богино	их
ER сүлжээ	Poissonian	богино	жижиг
Загвар сүлжээ	Custom, can be broad	богино	жижиг
Watts & Strogatz сүлжээ	Poissonian	богино	их
Barabasi Albert (Scale-Free) сүлжээ	Power-Law	богино	Жижиг

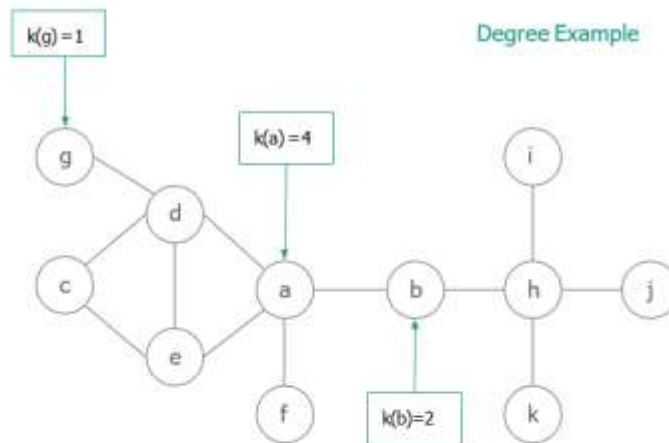
Зураг 22. Сүлжээнүүдийн үзүүлэлтүүдийн харьцуулалт

Tuyatsetseg Badarch, "Data communications and computer networking", third edition, 2014.

Зангилааны ач холбогдолыг үнэлэх

Бид зангилаануудын хувьд тухайн сүлжээндээ хэр чухал болохыг нь хэмжихдээ centrality хэмжүүр ашигладаг.

- Зарим centrality хэмжүүрүүд нь ойлгох тайлбарлахад энгийн амар хялбар байдаг. Мөн centrality хэмжүүрүүд нь граф дээр машин сургалтыг хэрэгжүүлэх зангилааны шинж чанараар (node features) ашиглагддаг.



Зураг 23. Сүлжээнүүдийн орой буюу зангилааны үнэлгээ

k = number of links

$$A_{i,j} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected,} \\ 0 & \text{otherwise} \end{cases}$$

$$k_i = \sum_{j=1}^n A_{i,j}$$

Сүлжээний зангилаа, графын орой

Зангилаа бүр өөрийн гэсэн төвлөрлийн зэргийн скортой. Зангилаа бүр өөрийн скороо хөрш рүүгээ илгээдэг. Зангилаа бүрийн бүх скорууд нь дараах томъёотой.

Эйгийн вектор (x) болон лямбдагийн утга (λ)-ийн хослол нь дараах харьцаагаар тодорхойлогдоно.

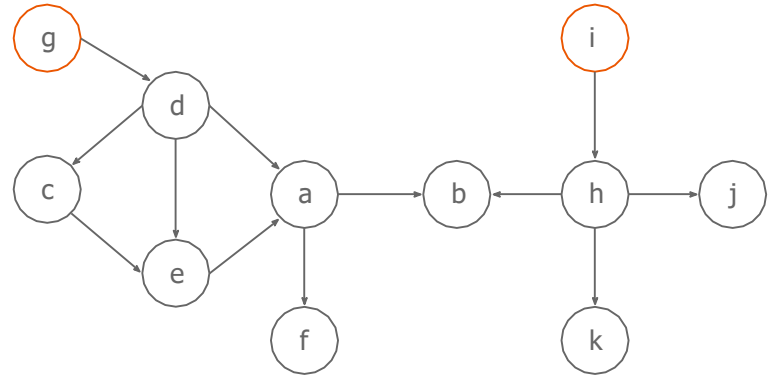
$$Ax = \lambda x$$

- n хэмжээтэй x вектор

- Ax нь хөршүүдээсээ хүлээж авсан нийт оноонуудын нийлбэр бүхий ижил хэмжээтэй шинэ вектор

The PageRank-ийн тооцооллын алгоритм нь Random Walk - процессоор тайлбарлагдана.

Санамсаргүй байдлаар дараагийн алхамд i оройд хүрэх магадлал нь зөвхөн одоогийн орой j болон транзишн магадлалаас хамааралтай. $j \rightarrow i$ нь стохастик матрицаар тодорхойлогдоно.



Зураг 24. Сүлжээнүүдийн орой буюу зангилаануудын графын орой

- Eigenvector centrality
- PageRank
- Katz centrality

Орой ойрхон шигүү сүлжээ байх чанарууд

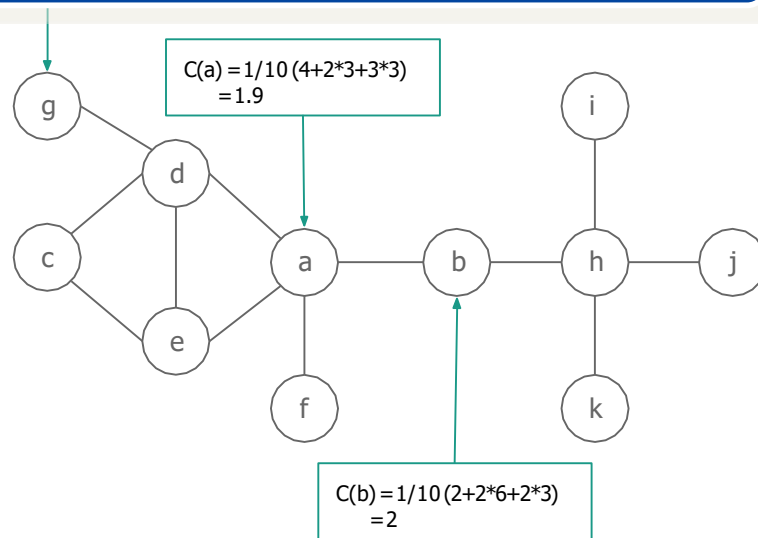
Closeness Centrality

тухайн оройгоор бусад бүх оройнуудад хүрэх богино замуудын уртын дундаж утга

Harmonic Centrality : Өгөгдсөн оройгоос бусад бүх оройнууд руу хүрэх зайны гарминик буюу зохицолт дундаж

Betweenness Centrality

Тухайн зангилаагаар дамжин явах боломжтой богино замуудын тоо



Closeness Formula

$$C_d(i) = \frac{n-1}{\sum_{d_{ij} < \infty} d_{ij}}$$

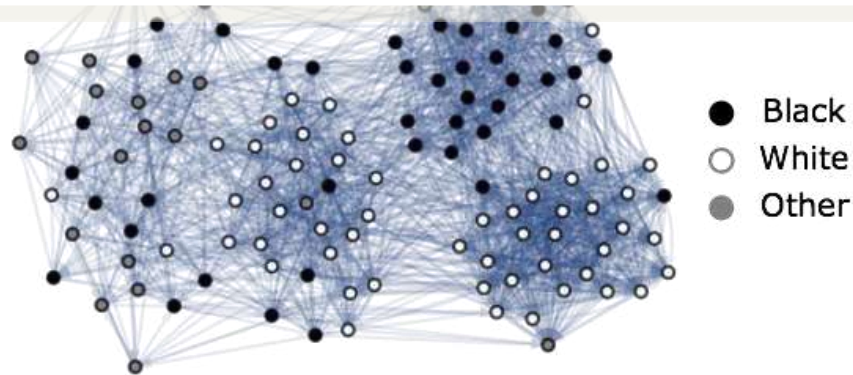
Зураг 25. Сүлжээнүүдийн орой буюу зангилаануудын ойрхон байх чанар

Сүлжээнд байх ижил чанар

Homophily Сүлжээнд байх ижил шинж чанар буюу атрибут бүхий оройнууд нь хоорондоо холбогдох магадлал нь тооцоолж байснаас өндөр байх хандлагатай.

Disassortative mixing: Contrary of homophily: өөр төрлийн буюу ижил бус оройнууд хоорондоо холбогддог хандлагатай сүлжээ ч байдаг жишээлбэл (e.g., sexual networks, predator-prey)

Холбоосын хүчин чадал & Уян хатан байдал



Examples of Vertex properties

age, gender, nationality,
political beliefs, socioeconomic status,
obesity, ...

Homophily can be a **link creation mechanism** or **consequence of social influence** (and it is difficult to distinguish)

Зураг 26. Сүлжээнүүдийн орой буюу зангилаануудын ижил байх чанар

Графын бүлэглэл үүсгэх

- *Граф дахь бүлэг гэж юу вэ?
- *Статик сүлжээн дэхь бүлгүүд
- *Бүлэг илрүүлэх алгоритмуудыг үнэлэх

Сүлжээний хувьд ойр байрлалтай бөгөөд бусад бүлгүүдэд байгаа объектуудтай харьцуулахад харьцангуй богино зайд байх зүйлс/объектууд нь нэг бүлэгт багтана.

Комплекс сүлжээний хувьд ойлгоход төвөгтэй топологиудыг тодорхойлох нь Бүлэг илрүүлэх алгоритмуудын зорилго юм.

meso scale - мезо масштаб

Бүлгийн шинж чанар нь ямар топтоги бүтэцтэй сүлжээ болохоос хамаарч өөр өөр аргуудыг ашиглаж болно.

- Чиглэлт/Чиглэлт бус граф
- Жинтэй/Жингүй граф
- Олон хэмжээст граф гэх мэт

Графын хувьд бусад олонлогт багтсан оройнуудыг бодвол илүү өөр хоорондоо нягт холбогдсон оройнуудын олонлог нь нэг бүлгийг үүсгэнэ.

- Топологи бүтцээс нь хамааруулаад оройнуудыг кластер болгох

Major Problems:

- Бүлэг илрүүлэлт нь ill posed буюу нэгээс олон хариулт олдох боломжтой бодлого юм. Өөрөөр хэлбэл, асуудлын шийдэл нь үндсэн өгөгдөл буюу граф ямар байгаагаас хамаарч өөр өөр бүлгүүд илэрч болох хандлагатай гэсэн үг.
- Олон ялгаатай аргууд ашиглагддаг
- Агуулгаас хамааралтай

Жижиг, энгийн сүлжээний хувьд бүлгүүдэд хуваах нь бүтцээс нь хараад амар хялбараар тодорхойлж болох зүйл юм. Гэвч комплекс буюу хэмжээний хувьд төвөгтэй их сүлжээн дээр энэ нь тийм ч амар биш.

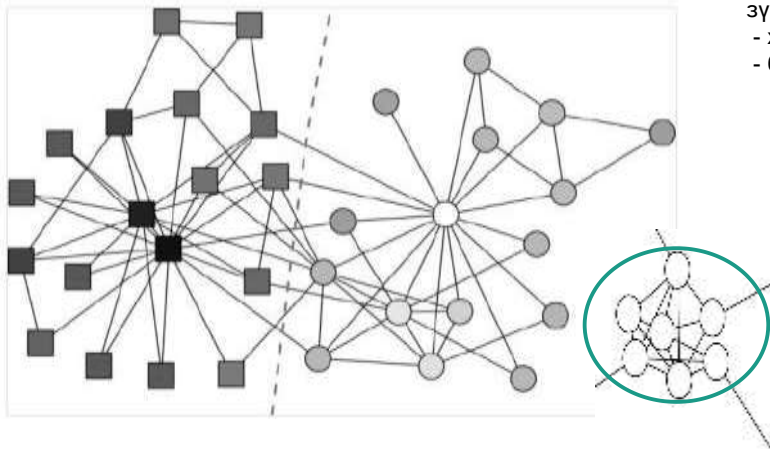
Tuyatsetseg Badarch, "Data communications and computer networking", third edition, 2014.

S. Fortunato. *Community detection in graphs*. Physics Reports 486 (2010).

Графын бүлэглэл үүсгэх

Ихэнх бодит сүлжээнүүд нүдээр харах, шинжилгээ хийхэд төвөгтэй, боломжгүй байдаг. Бид шууд хараад бүлгүүдийг тодорхойлж хэлж чадахгүй.

Тэгэхээр бидэнд бүлэг илрүүлэх процессыг автоматаар хийх арга алгоритмууд хэрэгтэй.



Зураг 27. Сүлжээнүүдийн орой буюу зангилаануудын бүлэг үүсгэх

T1: Бүлэг илрүүлэхийн тулд бүхэл сүлжээгээр хэрэн хэсэх хэрэгтэй.

T2: Нэг бүлэг нь нэг холбогдсон дэд граф болно.

T3: Бүлгүүд нь тухайн сүлжээн дэх локаль нягтралыг үүсгэдэг буюу бүлэг дотор буй хөрш оройнуудтайгаа нягт холбогдсон байдаг.

Бүлэг илрүүлэх алгоритмуудыг бид хэд хэд ангилж болох ба ингэхдээ дараах зүйлсийг харгалзан үздэг. Үүнд,

- хэр жижиг хэмжээний бүтэцүүдийг хайдгаар нь
- бүлгүүдийг илрүүлж арга замаар нь

Бүлгүүд нь нягт холбогдсон оройнуудын олонлог.

Бүлэг санамсаргүй үүсгэсэн графаас илүү олон тодорхой тооны ирмэгүүдтэй байх ёстой.

Сүлжээний нягтралыг тооцоолох

Нягтрал тооцох

Хэрхэн нягтралыг тооцох бэ?

- үнэлгээний функц нь бүлгийн нягтралыг тооцоолох ба уг функцын утгыг аль болох их байлгахыг хичээнэ.

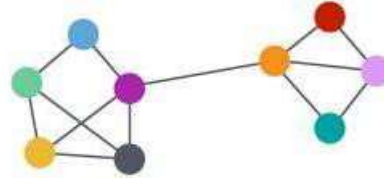
Modularity [-1, 1]

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

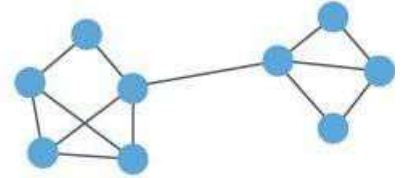
Таамагласан
нягтрал

1 if i, j нэг коммунити, 0
өөр

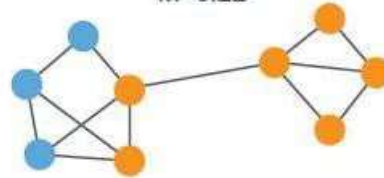
Нягтралын коэф
M=0.12



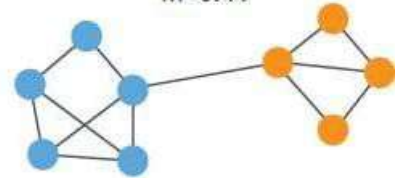
Нэг коммунити
M=0



хуваагдал | хуваагдал
M=0.22



хуваагдал | хуваагдал
M=0.41



Зураг 28. Сүлжээнүүдийн орой буюу зангилаануудын нягтрал олох

Girvan-Newman загвар

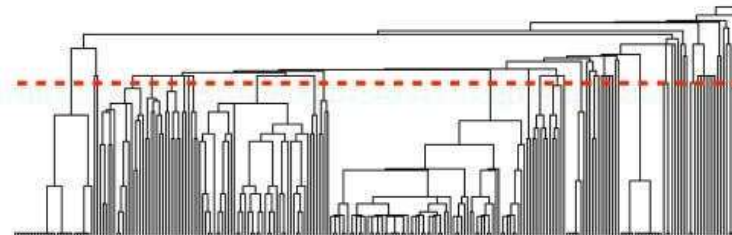
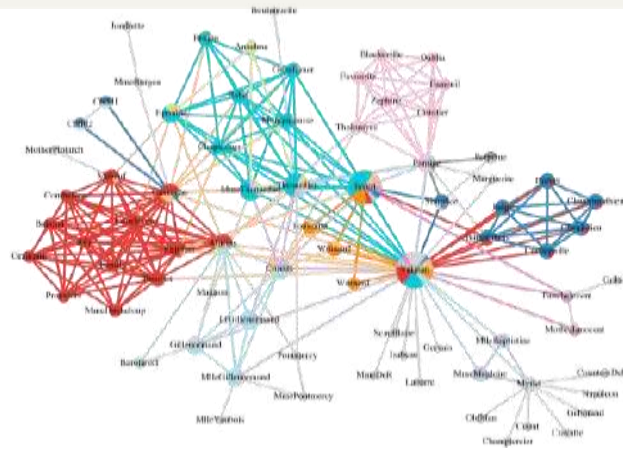
Алхамууд

1. Сүлжээн дэх бүх ирмэгүүдийн хувьд betweenness бодно.
2. Хамгийн өндөр betweenness centrality ирмэг(үүд)-ийг устгана.
3. Дахиад үлдсэн бүх ирмэгүүдийн хувьд betweenness-ийг бодно.
4. Ингээд 2 болон 3 алхамуудыг ирмэг үлдэхгүй болтол давтана.

Нягтрал ба Холбогч ирмэг

Энэ 2 тодорхойлолт төсөөтэй юм шиг санагдаж байна уу. Ижил ойлголт уу?

- Зарим тохиолдолд тийм
- Нягт, шигүү сүлжээнд холбогч ирмэгүүд нь тодорхой биш
- Харин сийрэг сүлжээнүүдэд холбогч ирмэгүүдийг ирмүүлэхэд амар боловч, нягтрлын хэмжээнд шууд нөлөөлнө.



Зураг 29. Сүлжээнүүдийн орой буюу зангилаануудын загвар

Зайн онцлог

Бүлгүүд нь ижил төсөөд шинж чанар (features)-уудтай байдаг тул үүнийг нь ашиглаад бүлэг илрүүлж болно.

Сонгогдсон оройн шинж чанар дээр үндсэнлэн зайн хэмжүүр тодорхойлогдоно.

Бүлэг доторх оролцогчид нь бусад бүлэг дэх оролцогчидыг бодвол өөр хоорондоо илүү төсөөтэй байдаг.

Percolation - Нэвчилт, нэвтрэлт

Бүлгүүд нь хамт нэг группэд багтах оройн олонлог бөгөөд эдгээр орой нь ижил шинжтэй, үйл ажиллагаа эсвэл мэдээлэл бүхий байх боломжтой.

Ихэвчлэн нэвчилтийн аргууд нь ямар нэгэн тодорхой үнэлгээний функцээр оптиминзац хийгддэггүй.

Label Propagation

1. Орой бүр нь дахин давтагдахгүй шошготой буюу id -тай.
2. Эхний алхам нь орой бүр дээр уг орой нь шошгоо өөрчлөх магадлал бодох ба энэ нь
3. Дараагийн алхам бүрт орой бүр хөрш оройтой ижил болгох боломжтой.
4. Энэ алхамуудыг бүх оройн шошго өөрчлөгдөж дуусах хүртэл хийнэ.

Algorithms in this family:

- Label Propagation
- Demon, Angel

Demon/ Angel

- Локаль хүрээнд, орой бүр өөрийнхөө бүлгийг тодорхойлох боломжтой.
- Глобал хүрээнд, бүлгүүдийн давхцал үүснэ.



Clustering кластерлалт

- Оройн шинжүүдийн төрөл ангилал нь зөвхөн хөрш оройнуудад ижил байна гэсэн үг биш юм.



Зураг 30. Сүлжээнүүдийн орой буюу зангилаануудын зайн онцлог

Графын бүтэц

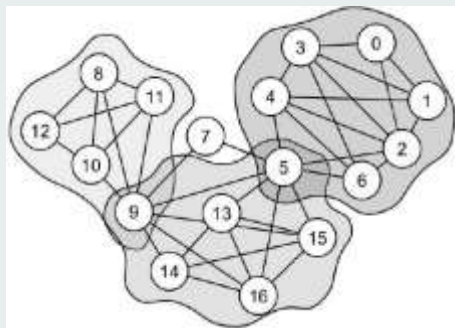
Оройнуудын олонлог буюу бүлгүүд нь тухайн оройнуудын холбосон олон тооны холбоос үүсгэдэг бас тодорхой топологуудаар тодорхой-логдох боломжтой.

Сүлжээн дэхь тодорхой хэв загвар, топлоги, паттерныг тодорхойлох

k-Cliques алгоритм

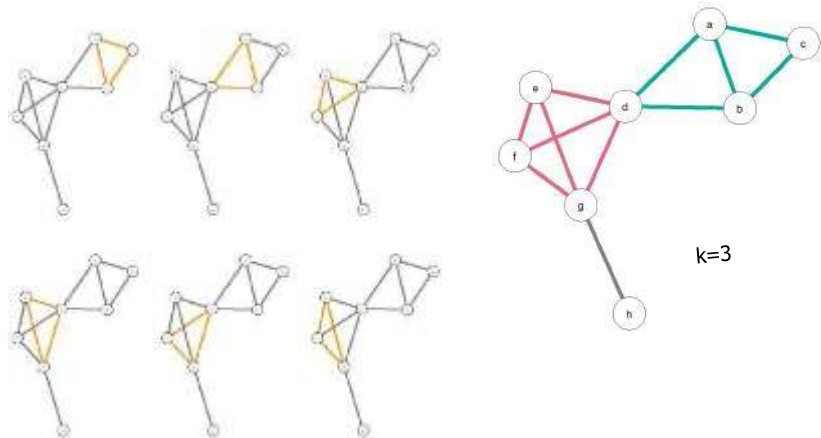
Хамгийн алдартай алгоритм бол k-cliques
- нягтралаас өөрөөр тодорхойлогдоно.

7 гэсэн оройн хувьд {9,5,7} гэсэн гурвалжины 1 орой б боловч ямар бүлэгт харъяалагдаагүй байна.



Алгоритм:
- k-cliques, ...

1. k-cliques тодорхойлох.
k-clique бол k орой бүгд өөр хоорондоо холбогдсон байна.
2. Холбогдсон хөрш k-cliques нэг бүлэгт орно.
k-cliques дэхь k-1 орой нь ижил бол хөрш clique буюу 1 бүлэгт багтана.



Зураг 31. Сүлжээнүүдийн орой буюу зангилаануудын алгоритм

Параметруудийг үнэлэх

Bipartite & Directed Networks

Урьд нь бид сүлжээг энгийн, чиглэлгүй, жингүй граф дээр авч үзсэн. Ялгаатай төрөл зүйлээр нь бүлэгт хувааж болно.

Жишээ:

- Antichains, Sibiliary Communities (DAG)
- One-to-One, Many-to-One (bipartite)

Attributed Networks

Орой болон ирмэгүүдийн нь өөрийн шинж чанарт бүхий нэмэлт атрибутууд бүхий семантик давхаргатай байж болно.

Жишээлбэл, орой нь хүний илтгэж байгаа бол тухай хүний нас, яс үндэс, боловсрол ... гэх мэт

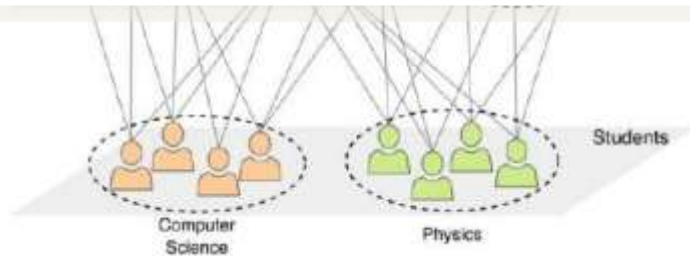
Тэгвэл графын топологи бүтэц болон семантик утга агуулгаар нь хоёулангаар нь сегменташн хийж утга учиртай хувааж болно.

Internal Evaluation

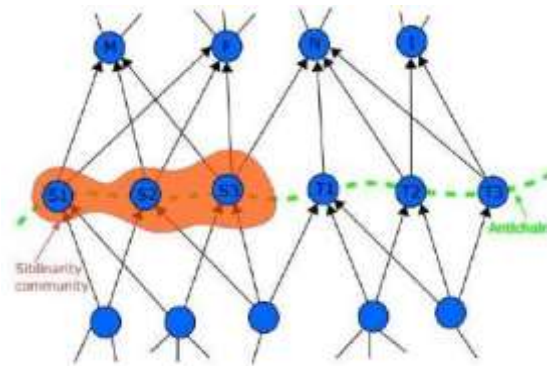
Хэд хэдэн фитнес функц ашиглан хуваагдсан бүлгүүдийн чанарыг Ихэвчлэн сайн хуваалт хийсэн бол уг функцуудын утга их гардаг.

Зураг 32. Сүлжээнүүдийн орой буюу зангилаануудын үзүүлэлтүүдийг үнэлэх

Tuyatsetseg Badarch, "Data communications and computer networking", third edition, 2014.



Taguchi, Hibiki, and Tsuyoshi Murata. "BiMLPA :Community Detection in Bipartite Networks by Multi-Label Propagation." NetSciX (2020).



Vasiliauskaite, Vaiva, and Tim S. Evans. "Making Communities Show Respect for Order." arXiv preprint arXiv:1908.11818 (2019).

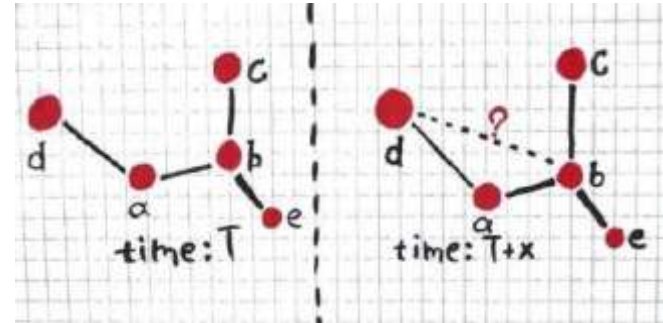
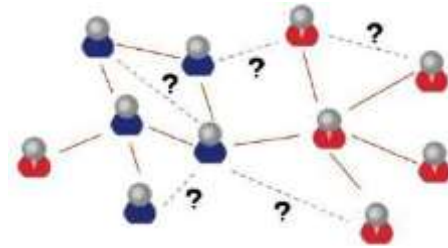
Шугам, холбоосын таамагал

Link Prediction

- Сүлжээний өөрчлөлтийг таамаглах нь
 - * Удирдлагагүй сургалтыг аргууд
 - * Удирдлагатай сургалтыг аргууд
 - * Үнэлгээ

Сүлжээний өөрчлөлтийг хэрхэн ойлгох вэ?

Нэгэн сүлжээний t хугацаан дахь снаппшот өгөгдсөн бол $(t, t+1)$ интервалын турш тухайн сүлжээнд нэмэгдэх ирмэгүүдийг хэрхэн нарийвчлал сайтай таамаглах вэ?



Террорист сүлжээг хянах - террористуудын хооронд боломжит харилцаануудыг багасгах эсвэл орхигдсон линкийг таамаглах

Байгуулага дах хамтран ажиллагсадыг олоход/хооронд нь холбоход туслах

ҮНайзуудыг таамаглах

Ирээдүйд үүсэх боломжтой холбоосыг тухайн сүлжээнд олохдоо холбоосуудыг таамаглах аргыг ашиглана.

Эсвэл, бүрэн мэдээлэл байхгүй тохиолдолд орхигдсон холбоосуудыг таамаглаж болно.

Зураг 33. Сүлжээнүүдийн орой буюу зангилаануудын шугам, холбоосын таамаглал

Tuyatsetseg Badarch, "Data communications and computer networking", third edition, 2014.

Liben-Nowell, David, and Jon Kleinberg. "The link-prediction problem for social networks." *Journal of the American society for information science and technology* 58.7 (2007): 1019-1031.

Шугам, холбоосын таамаглал

Link Prediction

Өгөгдсөн граф $G=(V,E)$ бол уг сүлжээний бүх боломжит ирмэгүүдийг таамаглахад $O(|V|^2)$ байна. Бодит сүлжээнүүд нь сийрэг байх хандлагатай.

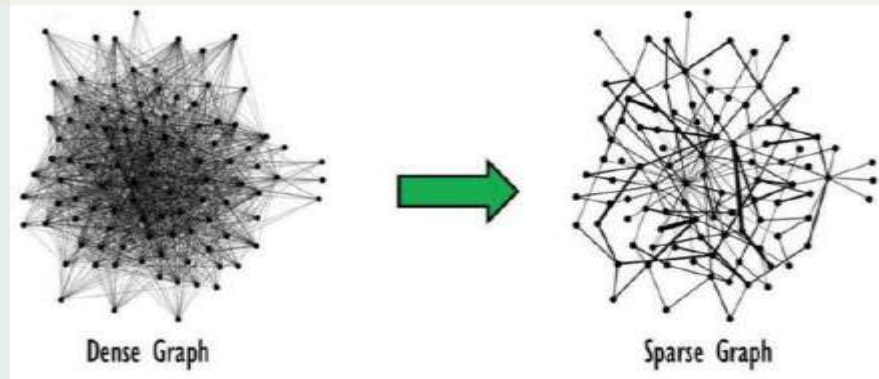


Бидний хэзээ ч үүсэхгүй ирмэгүүдийг таамаглах магадлал нь False Positive таамаглалын алдаа!!

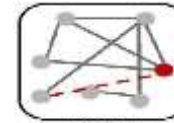
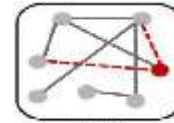
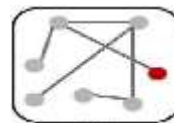
Эрдэмтэд сүлжээнд илүү ойр бол (дундын хамрагчид олонтой) -> ирээдүйд тэд хамтран ажиллах боломжтой.

энэ сэдлээс үндэслэн нарийвчлах, ойлгохдоо proximity гэх хэмжүүрийг таамаглалын нарийвчлалыг олоход ашигладаг.

Зураг 34. Сүлжээнүүдийн шугам, холбоосын таамаглал



1. t хугацаанд оролтын нэг граф G гэж үзье
2. (u,v) оройнуудыг холбосон бүх боломжит оройнуудыг авч үзье
3. $score(u, v)$ гэх функцээр холбоосыг тодорхойлъя.
4. Дээрх $score$ -уудыг буурахаар эрэмбэлж ирмэгүүдийн жагсаалтыг үүсгэнэ.
5. $t+1$ эгшинд граф дахь ирмэгүүдийг таамаглахдаа дээрх жагсаалтаа ашиглана. $score$ нь proximity-ийн хэмжүүр.

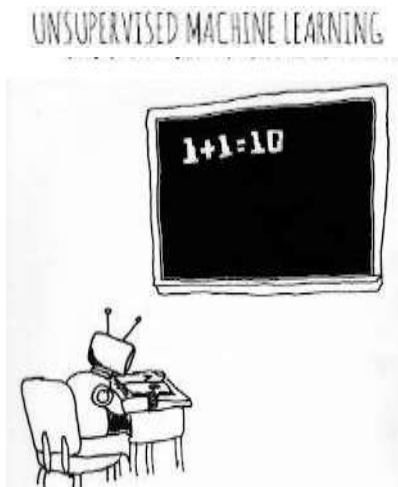


Аргууд

Link Prediction

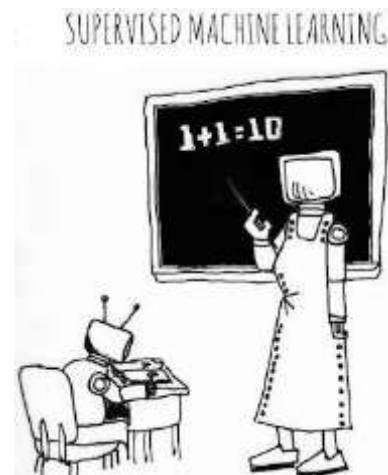
Unsupervised

Тухайн сүлжээтэй холбоогүйгээр дөхүү хэмжүүрүүдийн олонлогыг ашиглан тодорхойлно.



Supervised

Таамаглалын үр дүн нарийвчлалыг сайжруулахын тулд тухайн сүлжээнээс мэдлэг гаргаж авдаг.



Зураг 35. Сүлжээнүүдийн орой буюу зангилаануудын таамаглалын аргууд

Удирдлаггүй арга

Удирдлагагүй холбоосыг таамаглах арга

Хэв шинжийг олохдоо зөвхөн оролтын өгөгдлийг ашигладаг аргыг хэлдэг.

Өгөгдлөөс яг юу хайхаа мэдэхгүй байгаа үед энэ арга тохиромжтой.

Түүхий өгөгдлийг ойлгохын тулд ихэвчлэн ашигладаг.

Энэхүү удирдлаггүй арга нь системийн агуулгыг (үзүүлэх шинжилгээ) гэдэг техник дээр суурилна.

Neighborhood measures

- Common Neighbors, Adamic Adar, Jaccard, Preferential Attachment

Path-based measures

- Graph distance, Katz

Ranking

- Sim Rank, Hitting time, Page Rank

Хөршийн хэмжүүрүүд

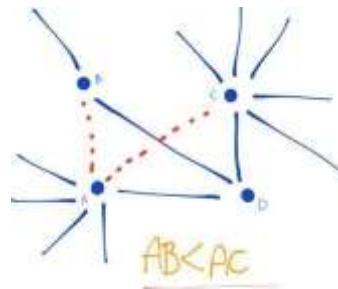
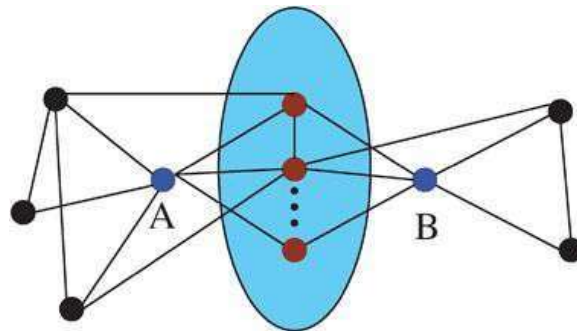
Unsupervised Link Prediction

Найзлахын тулд хэчнээн дундын найз байх ёстой вэ?

Дундын хөрш оройнууд:
Бид олон дундын найзтай бол илүү
найзлах хандлагатай.

$$\text{score}(u, v) = |\Gamma(u) \cap \Gamma(v)| \quad \text{score}(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

Найзууд болохын тулд бид хэдэн шэйр хийх вэ



Зураг 36. Сүлжээнүүдийн орой буюу зангилаануудын хөршийн хэмжүүрүүд

ҮНЭЛГЭЭ

Өгөгдсөн p таамаглагч байвал уг таамаглагчийг сайн эсэхийг хэрхэн мэдэх вэ?

First Step:

p нь санамсаргүй таамаглагчаас сайн үзүүлэлттэй байх

Random Predictor

Сүлжээн дэх ирмэг бүр ижил магадлалтай.

$$\text{performance}(p) = \frac{TP}{TP + FP}$$

$$\text{ratio} = \frac{\text{performance}(p)}{\text{performance}(\text{prandom})} = \frac{\text{performance}(p)}{\frac{V(-(\sqrt{V}-1))}{2} - |E_{\text{old}}|}$$

1. Сүлжээ ихэвчлэн сийрэг
2. Cold Start асуудал
Хангалттай өгөгдөл байхгүй бол юу болох вэ?
3. False Positive буюу буруу байх магадлал их
4. Энгийн аргууд нь хэтэрхий энгийн.
5. Комплекц аргууд нь хэтэрхий өртөг өндөр.

Динамик сүлжээний бүлэг илрүүлэлт

Цаг хугацаа өнгөрөхөд сүлжээн дэх

- зангилаанууд үүсч бас алга болно
- холбоос ирмэгүүд илэрч бас устана

... тэгвэл тухайн сүлжээн дэх бүлгүүд мөн өөрчлөгдөнө!

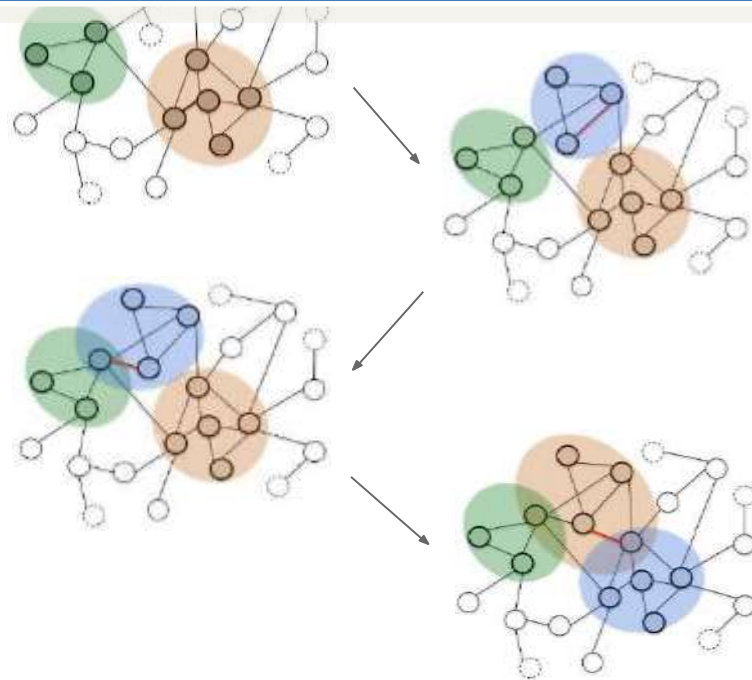
Бүлгүүдийн амьдралын цикл

Шинэ орой болон ирмэгүүд үүсч, устах нь сүлжээний эмх замбраагүй байдлыг бий болгодог.

Бүлгүүд ч мөн уг өөрчлөлтөнд хүчтэй өртөнө.

- тодорхойлох & тохируулах
- давтагдах алгоритм ашиглах
- тогтонги байдлыг илрүүлэх

1. Сүлжээ бүрийг ашиглаж бүлгүүдийг олох
(статик алгоритм ашиглах)
2. Дараалсан сүлжээнүүд дээр бүлгүүдийг
жиших
3. Ялгаануудыг ашиглах

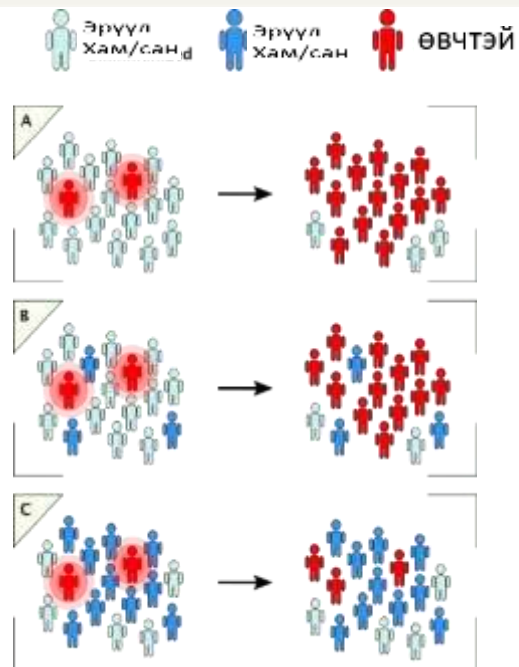
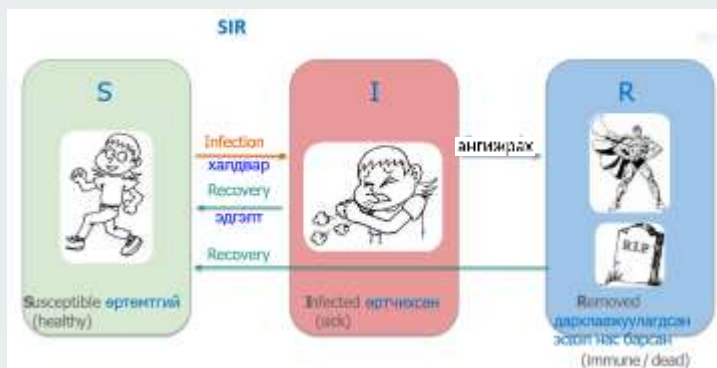


Cazabet, Remy, and Giulio Rossetti. "Challenges in community discovery on temporal networks." *Temporal Network Theory*. Springer, Cham, 2019. 181-197.

Тархалтын загварууд

Нэвтрэлтэнд шийдэлд суурилсан загварууд

Дундаж хэсгийн утгаар загварчлах



Зураг 39. Сүлжээнүүдийн нэвтрэлтийн чанар, тархалт



АШИГЛАСАН МАТЕРИАЛ

Tuyatsetseg Badarch, "Data communications and computer networking" , third edition, 2014.

Foundations of Computer Science, Behrouz A. Forouzan, Fourth Edition, Cengage Learning EMEA, 2018. Бүлгийн гарчиг “Chapter 6. Networks and Internet ”





АНХААРАЛ ХАНДУУЛСАНД БАЯРЛАЛАА