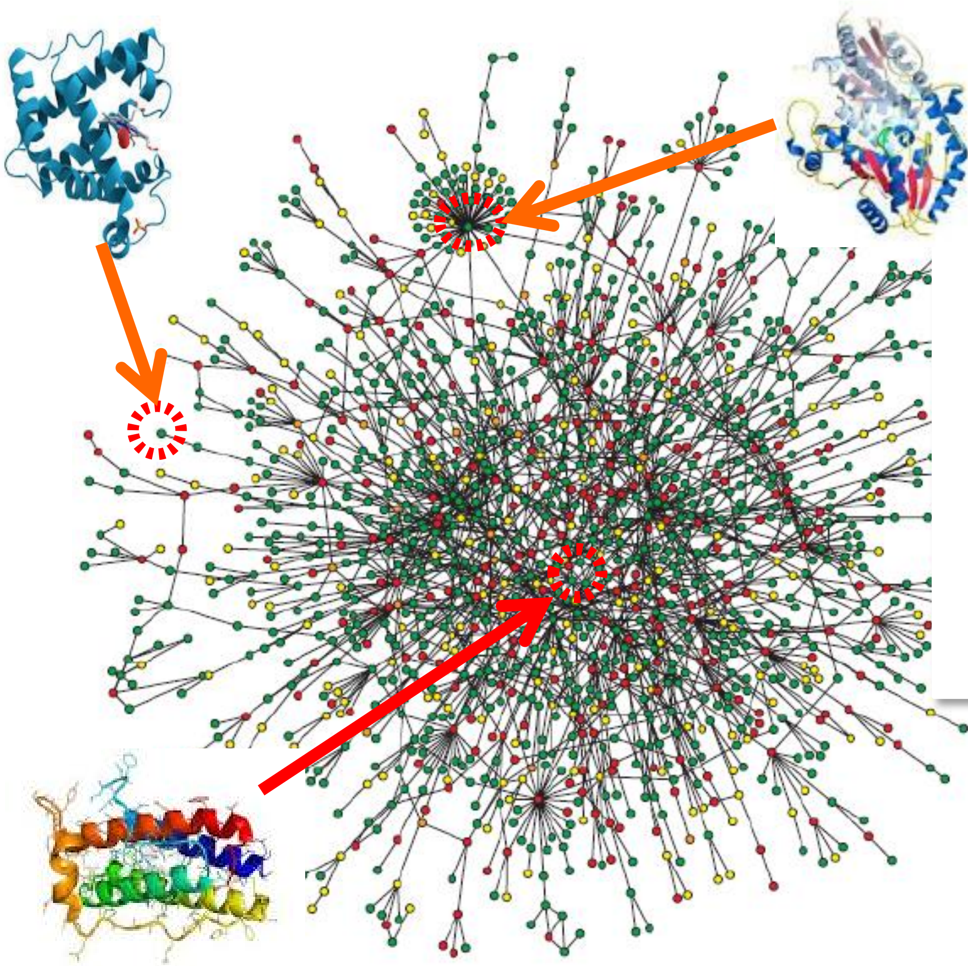
Centralidad

¿Qué tan *relevante* es un nodo de la red?

Topológia y rol biológico



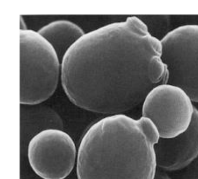
Redes Complejas

Degree
Eccentricity
Closeness
Betweenness
Bridging centrality
Eigenvalue centrality
Random walk centralities
….

Biologia

Essentiality
Phenotypic variability
Biological function
Cancer related
Disease Related
…

Protein-protein interaction in yeast *S. cerevisiae*, (N=1870 and L=2240).
Jeong et al., Nature, 411, 41 (2001).



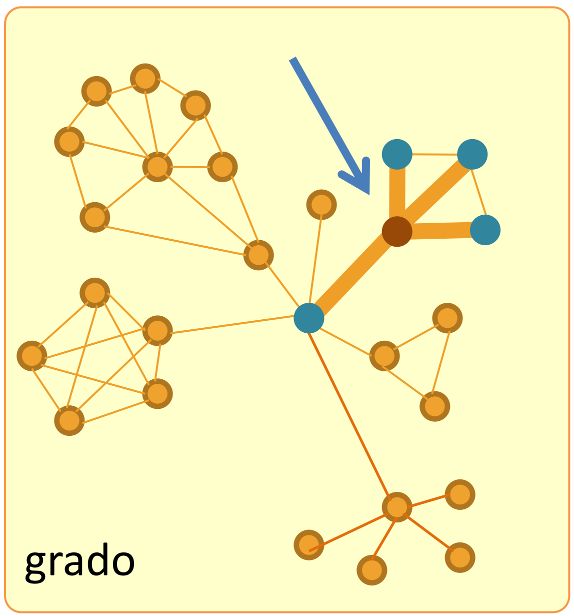
Propiedades topológicas: entorno local/global

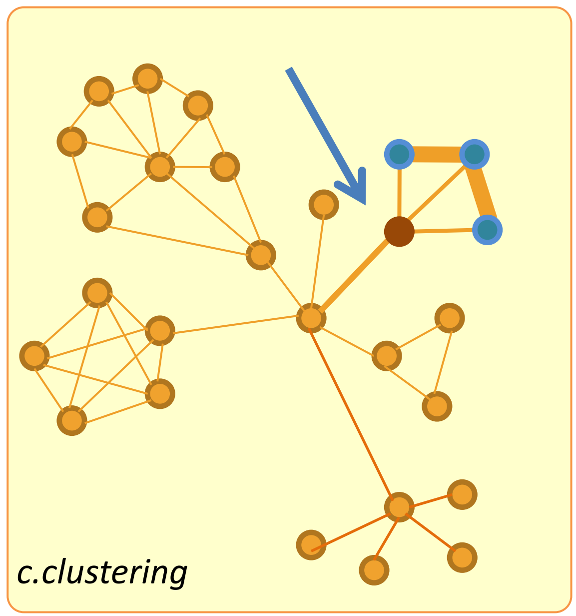
Propiedades Locales

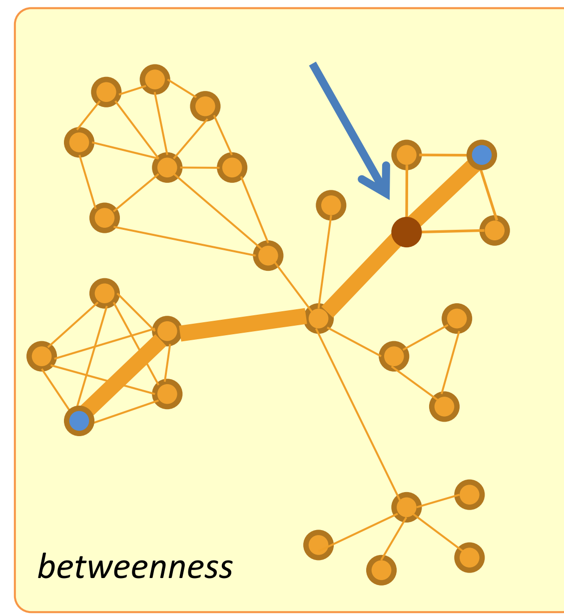
grado
coef. Clustering
centralidad ‘random walk’
…

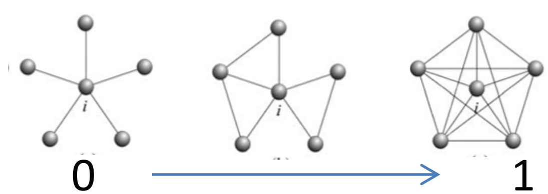
Propiedades Globales

betweenness
centralidad de autovalores adyacencia
cercania
…



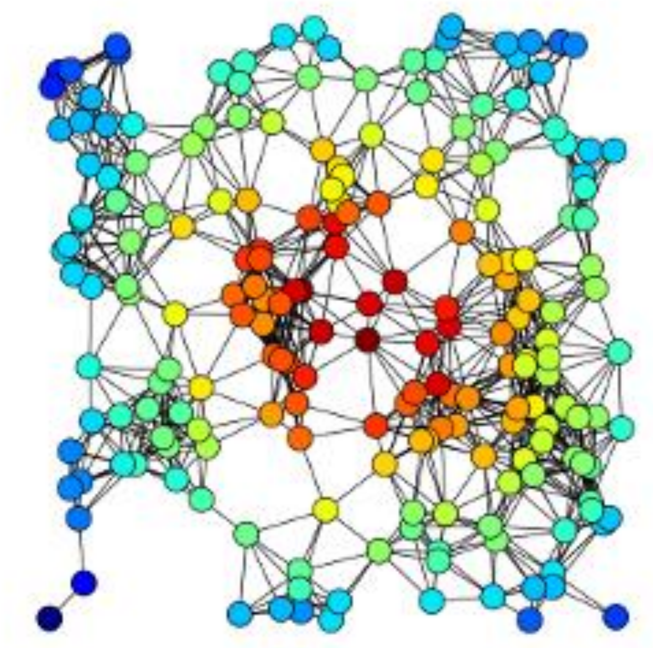


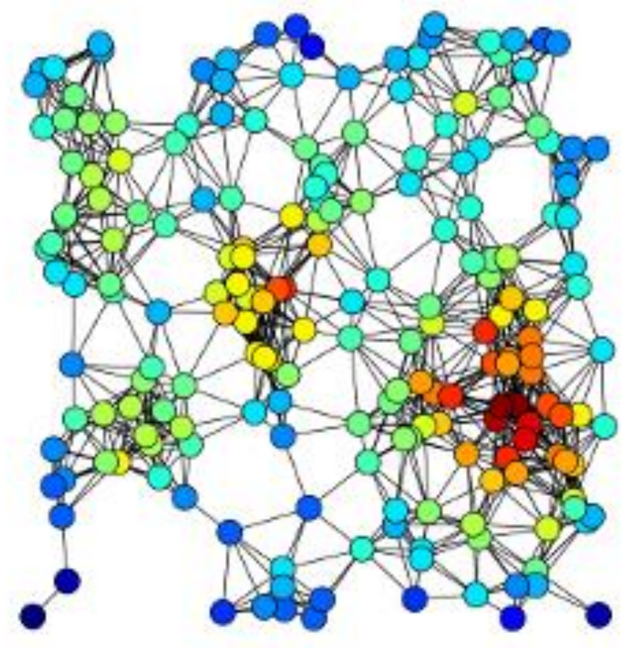




$$be(v) = \sum_{s \neq t \neq v} \frac{\sigma_{s,t}^{shortest-path}(v)}{\sigma_{s,t}^{shortest-path}}$$

Que significa ser **importante**?





Noción de **flujo** sobre la red

Noción de **coesividad** sobre la red

Dime qué modelas con tu red y te diré qué es ser importante

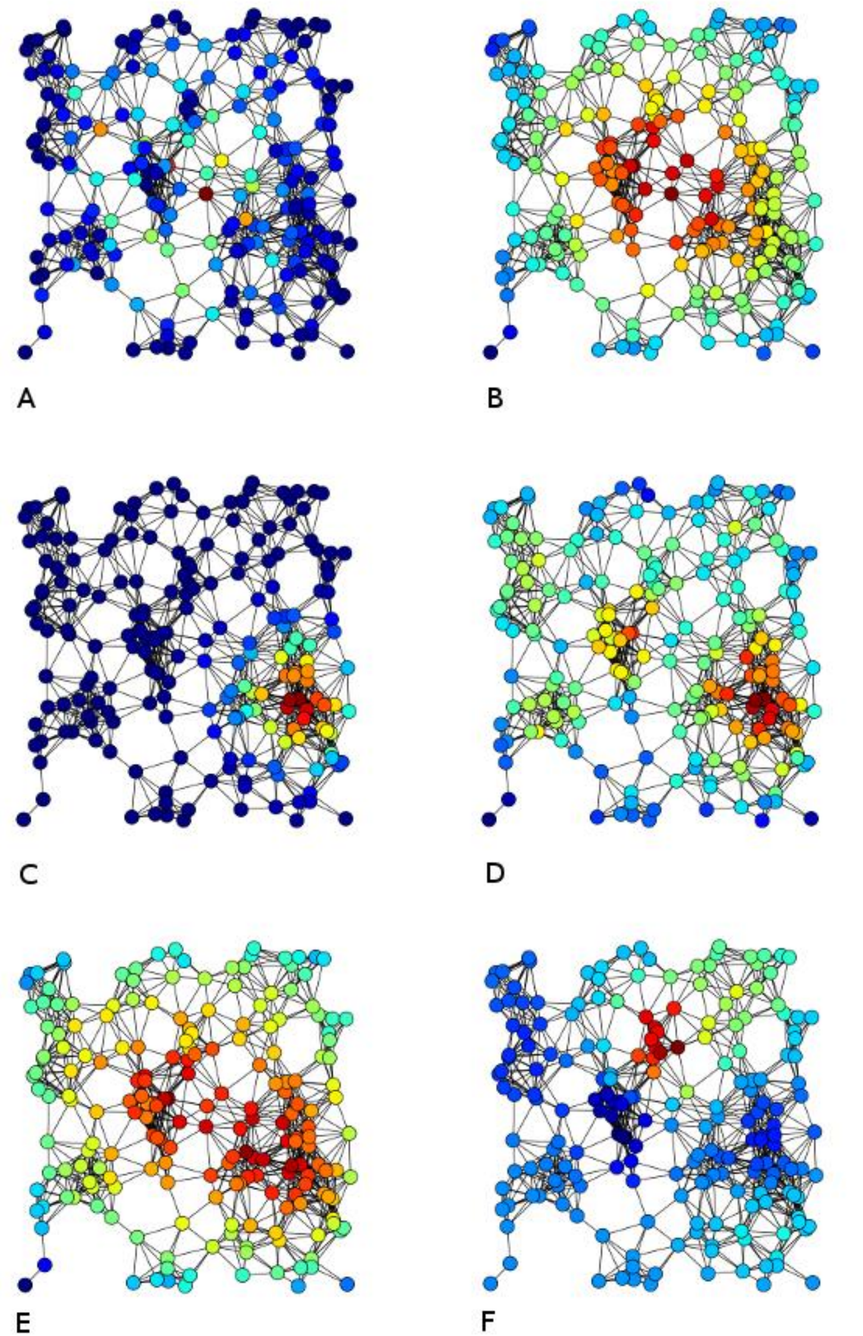
Centralidad

Dime qué modelas con tu red y te diré qué es ser importante

Noción de **flujo** sobre la red

Noción de **coesividad** sobre la red

a. intermediatez
b. cercanía
c. autovalor
d. grado
e. centralidad armónica
f. centralidad de Katz



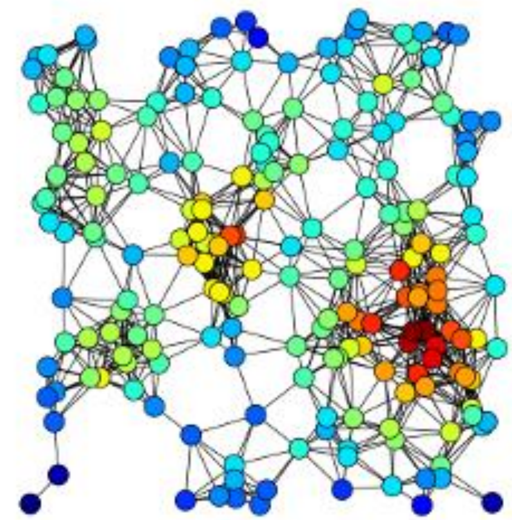
Centralidad y flujos

Dime qué modelas con tu red y te diré qué significa ser importante

- Flujo de items indivisibles que están en un único lugar a un dado tiempo. Su difusión es por transferencia. Puede haber casos en que:
  - Billetes, libros: se mueven por el grafo **sin restricciones** acerca de repetir enlaces o nodos ya visitados (i.e. *walks*)
  - Ropa usada, *re-gifting:* se mueven por el grafo **sin repetir enlaces** (i.e *trails*)
  - correo, gps: se mueven por el grafo desde un nodo origen hacia uno destino minimizando distancia recorrida (i.e. geodésicas).

- Flujo de entidades que pueden estar en varios lugares a la vez. Difusión por replicación.
  - Chismes: Transmisión boca-a-boca, en general no repite enlaces, pero si vértices (i.e. *trails*)
  - Campaña mailing: Transmisión por emisora (*broadcasting),* suele ser simultánea.
  - Ideas, creencias, modas, actitudes: influencia que se transmite boca-a-boca por replicación. Puede repetir enlaces (i.e. *walks*)
  - Procesos infecciosos, con inmunidad. Procede por replicación, pero sin revisitar vértices (i.e. *paths*).

Centralidad de **grado**



- Utiliza directamente el **grado** de un nodo como medida de su importancia

$$Cent(n_i) = k_i$$

- En términos de flujo:
  - caracteriza efectos de influencia inmediata.
  - razonable por ejemplo para aplicar en procesos de duplicación paralela (prob de recibir algo que esta distribuido aleatoriamente por la red es proporcional al nro de contactos) o de caminatas al azar.

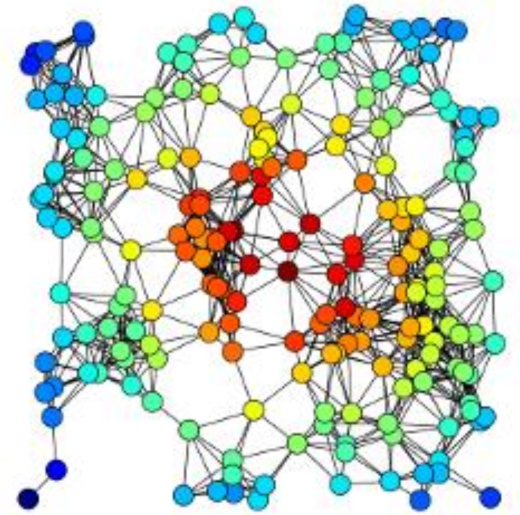
- En términos de cohesividad:
  - Hubs proveen atajos entre pares de nodos

- Asume linealidad: un nodo con el **doble** de vecinos que otro es **dos** veces mas importante

- Sólo utiliza información **local**

Centralidad de **cercanía**

- Asume que los nodos mas importantes son aquellos que pueden alcanzar fácilmente a más vecinos (están **cerca** de todos)



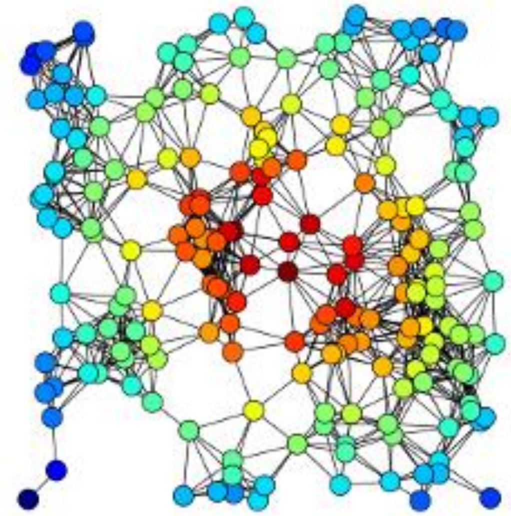
$$Closeness\ Centrality(n_i) = \left[\frac{\sum_{j\neq i} d_{geod}(n_i, n_j)}{N-1}\right]^{-1}$$

(sin vecinos) $0 \leq ClosensessCentrality \leq 1$ (vecino de todos)

- nodos que en promedio reciben flujo transmitido por la red rápidamente:
  - red de flujo de información: ubicaciones ventajosas
  - red de contactos sexuales : ubicaciones riesgosas
  - en gral para flujos a lo largo de geodésicas o por duplicación en paralelo
  - no seria un buen indice para identificar, por ejemplo, quién recibirá info antes si el proceso de transmisión es tipo rumor (no va por caminos mas cortos)

Centralidad de **cercanía**

- Asume que los nodos mas importantes son aquellos que pueden alcanzar fácilmente a más vecinos (están **cerca** de todos)



$$Closeness\ Centrality(n_i) = \left[\frac{\sum_{j \neq i} d_{geod}(n_i, n_j)}{N-1}\right]^{-1}$$

(sin vecinos) *0 ≤ ClosensessCentrality ≤ 1* (vecino de todos)

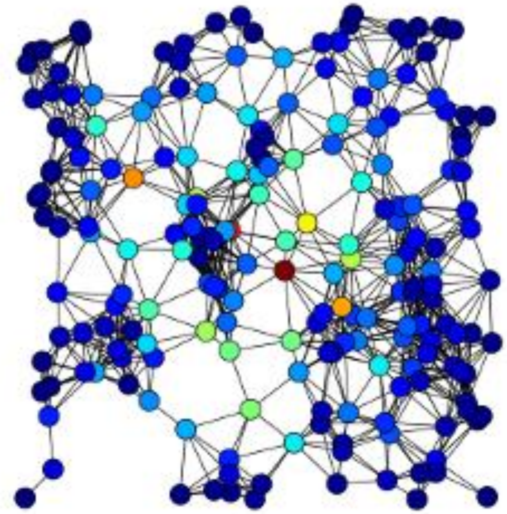
- En general tiene un rango dinamico chico (distancias geodésicas son chicas 1< d_{geod}< o(log N))
- No está bien definido para redes con más de una componente.
- Aún si calcularamos cada componente por separado no es trivial dar un ordenamiento global (centralidad en componentes más chicas tiende a ser mayor)

$$Harmonic\ Closeness\ Centrality(n_i) = \frac{1}{N-1}\sum_{j \neq i}\frac{1}{d_{geod}(n_i, n_j)}$$

- Bien definido aún para redes con más de una componente.
- Se le da más peso a entorno local

Centralidad de **intermediatez** (***betweenness***)

Si hay un intercambio global de mensajes entre **todos los pares** de nodos de la red y si la información se intecambia por **geodésicas**: el nodo por el que pasarán más mensajes es aquél por el que pasan la mayor cantidad de geodésicas de la red.



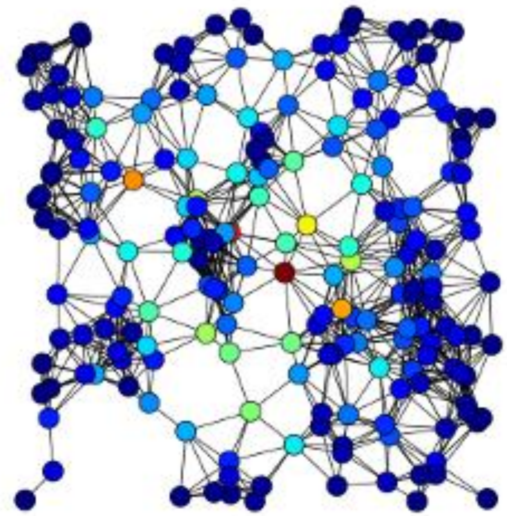
$$bet(n_i) = \sum_{r,s} \frac{x^{i}{}_{rs}}{x_{rs}}$$

← nro de geodésicas entre *r* y *s* que pasan por *i*

← nro de geodésicas entre *r* y *s*

Centralidad de **intermediatez** (***betweenness***)

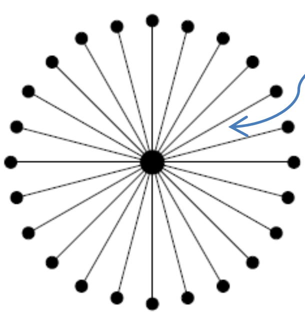
Si hay un intercambio global de mensajes entre **todos los pares** de nodos de la red y si la información se intecambia por **geodésicas**: el nodo por el que pasarán más mensajes es aquél por el que pasan la mayor cantidad de geodésicas de la red.



$$bet(n_i) = \sum_{r,s} \frac{x^{i}{}_{rs}}{x_{rs}}$$

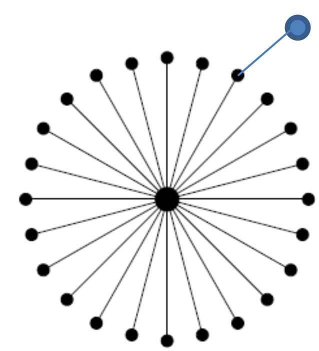
nro de geodésicas entre *r* y *s* que pasan por *i*

nro de geodésicas entre *r* y *s*



$$bet_{max} = n^2 - (n-1)$$

Participa en la geodesica de todos los pares menos en la de los n-1 perifericos consigo mismos

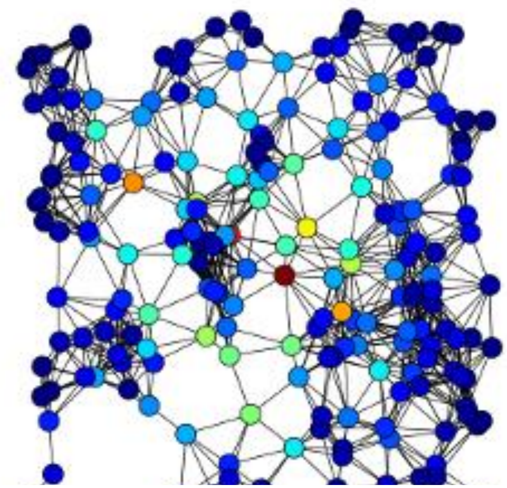


$$bet_{min} = 2n - 1$$

Participa en las geodesicas (ida y vuelta) del azul con el resto 2(n-1) mas la del azul consigo mismo.

Centralidad de **intermediatez** (***betweenness***)

Si hay un intercambio global de mensajes entre **todos los pares** de nodos de la red y si la información se intecambia por **geodésicas**: el nodo por el que pasarán más mensajes es aquél por el que pasan la mayor cantidad de geodésicas de la red.



$$bet(n_i) = \sum_{r,s} \frac{x^i{}_{rs}}{x_{rs}}$$

nro de geodésicas entre *r* y *s* que pasan por *i*

nro de geodésicas entre *r* y *s*

$$bet_{min} = 2n - 1$$

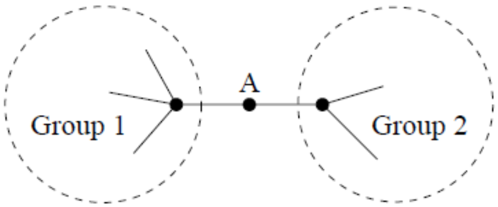
$$bet_{max} = n^2 - (n - 1)$$

$$\frac{bet_{max}}{bet_{min}} \sim \frac{1}{2} n$$

(amplio rango dinámico)

Es una medida global

Es una medida de intermediatez más que de conectividad



Centralidad de **autovalor**

Nodos importantes tienen **vecinos** importantes

- x_i (importancia de nodo-i) depende de x_j (importancia de nodo-j vecino). La importancia se *transmite* por la red.
- la manera más simple es asumir linealidad

$$x_i' = \sum_{j=1}^{N} a_{ij} x_j \qquad \boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x}$$

- el nivel de importancia se establece recursivamente en la red

$$\boldsymbol{x}'' = \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{A}\boldsymbol{x}$$

$$\vdots$$

Despues de ***s*** iteraciones: $\boldsymbol{x}(s) = \overbrace{\boldsymbol{A}\boldsymbol{A}\ldots.\boldsymbol{A}}^{\boldsymbol{s}}\,\boldsymbol{x}(0)$

$$\boldsymbol{x}(s) = \boldsymbol{A}^{\boldsymbol{s}}\boldsymbol{x}(0)$$

Centralidad de **autovalor**

Nodos importantes tienen **vecinos** importantes

- x_i (importancia de nodo-i) depende de x_j (importancia de nodo-j vecino). La importancia se *transmite* por la red.
- la manera más simple es asumir linealidad

$$x_i{}' = \sum_{j=1}^{N} a_{ij} x_j \qquad \boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x}$$

- el nivel de importancia se establece recursivamente en la red

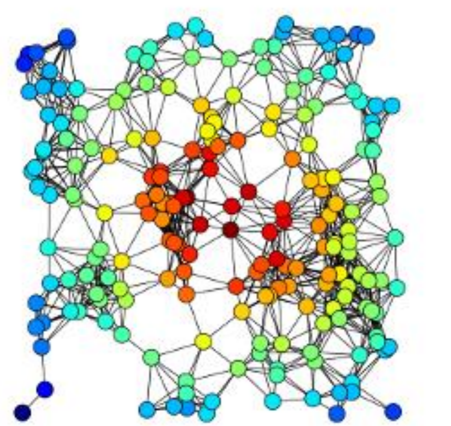
$$\boldsymbol{x}'' = \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{A}\boldsymbol{x}$$

$$\vdots$$

Despues de ***s*** iteraciones: $$\boldsymbol{x}(s) = \overbrace{\boldsymbol{A}\boldsymbol{A}\ldots.\boldsymbol{A}}^{\boldsymbol{s}}\,\boldsymbol{x}(0)$$

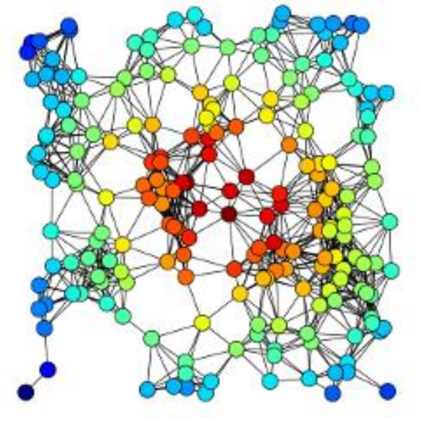
x solución autoconsitente $$\boldsymbol{x} = \boldsymbol{A}^{\boldsymbol{s}}\boldsymbol{x}$$

s>>1



x es un autovector de ***A***

Centralidad de **autovalor**



Busco $\boldsymbol{x}$ solución autoconsitente

$$\boldsymbol{x(s) = A^s x(0)}$$

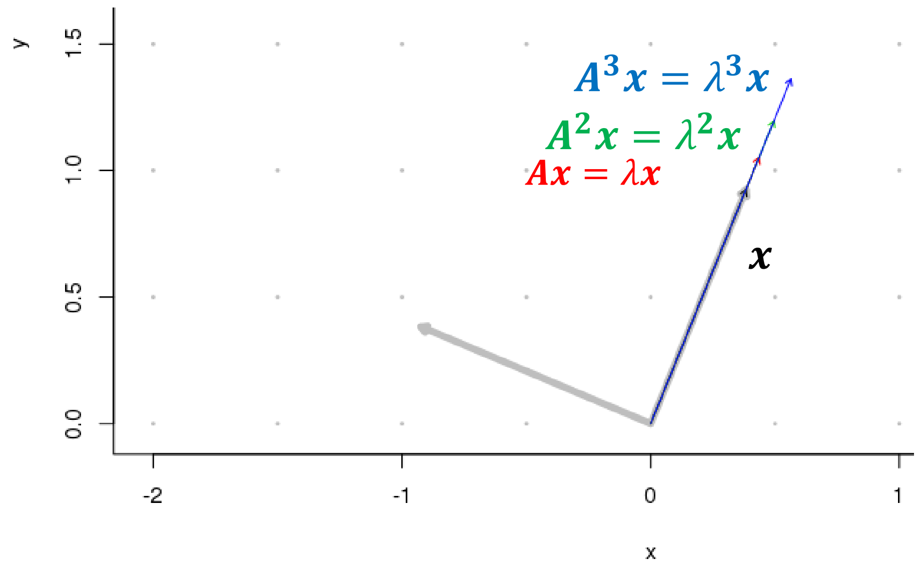
Si inicialmente arranco de un autovector: $\boldsymbol{x}(0)=\boldsymbol{v_k}$

$$\boldsymbol{x}(s) = \overbrace{\boldsymbol{AA\ldots.A}}^{\boldsymbol{s-1}}\,\boldsymbol{A\,v_k} = \lambda_k \overbrace{\boldsymbol{AA\ldots.A}}^{\boldsymbol{s-1}}\;\boldsymbol{v_k}$$

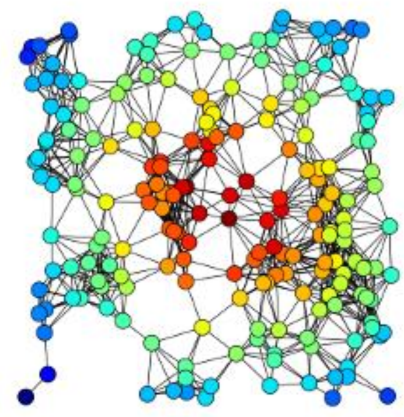
$$\boldsymbol{x}(s) = \lambda_k{}^{s}\,\boldsymbol{v}_k$$

Por lo que la **importancia relativa** de cada nodo se podría inferir de las componentes de $\boldsymbol{v_k}$

si $\boldsymbol{x}$ inicial fuera justo un **autovector**, por ejemplo $\boldsymbol{v_k}$, de la matriz de adyacencia $\boldsymbol{A}$

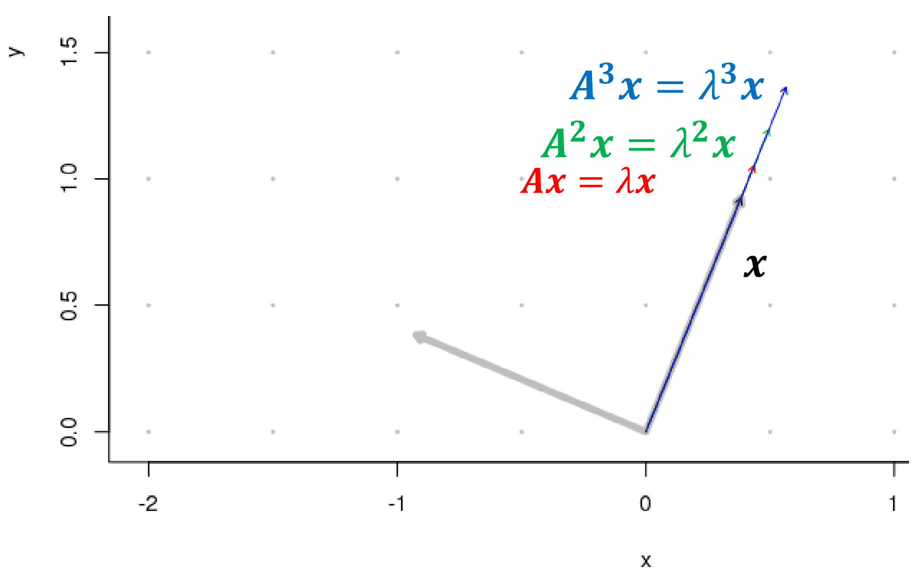


Centralidad de **autovalor**

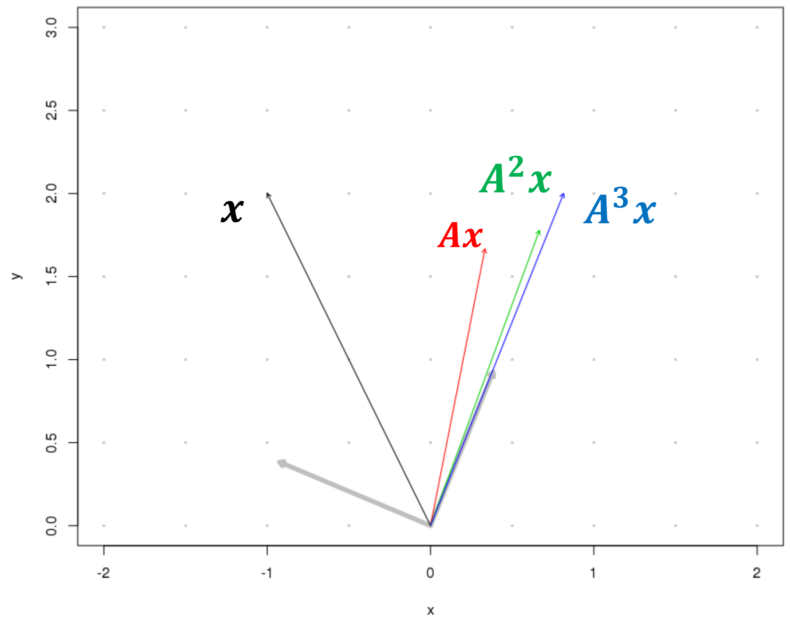


$$\boldsymbol{x}(s) = \boldsymbol{A}^s \boldsymbol{x}(0)$$

si ***x*** inicial fuera justo un **autovector**, por ejemplo $\boldsymbol{v}_k$, de la matriz de adyacencia ***A***

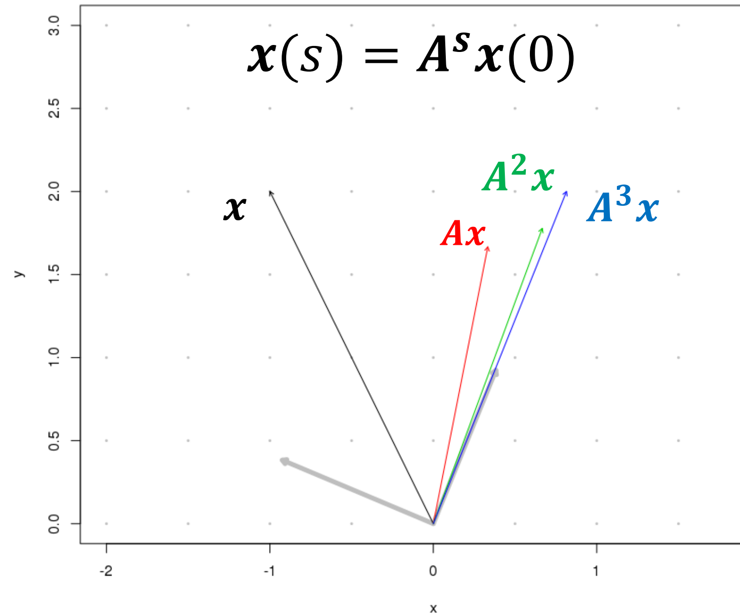


En el caso general ***x(0)*** no coincide con una dirección privilegiada asociada con ***A*** ....



Centralidad de **autovalor**

Para ver como una **matriz** transforma vectores es útil pensar en término de sus **autovectores** $\boldsymbol{v}_i$ (se transforman muy fácil: $\mathbf{A}\,\mathbf{v}_i=\lambda_i\,\mathbf{v}_i$). El vector $\boldsymbol{x}(0)$ puede ser escrito como combinación lineal de los $\{\mathbf{v}_i\}$



$$\boldsymbol{x}(0) = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots = \sum_{\boldsymbol{i}} c_i\boldsymbol{v}_{\boldsymbol{i}}$$

$$\boldsymbol{A}^s\boldsymbol{x}(\boldsymbol{0}) = \sum_{\boldsymbol{i}} c_i\boldsymbol{A}^s\boldsymbol{v}_{\boldsymbol{i}} = \sum_{\boldsymbol{i}} c_i{\lambda_i}^s\boldsymbol{v}_{\boldsymbol{i}} = {\lambda_*}^s\sum_{\boldsymbol{i}} c_i\left(\frac{\lambda_i}{\lambda_*}\right)^s\boldsymbol{v}_{\boldsymbol{i}}$$

mayor autovalor

$$\boldsymbol{x}(s) = {\lambda_*}^s\sum_{\boldsymbol{i}} c_i\,\overbrace{\left(\frac{\lambda_i}{\lambda_*}\right)}^{<1}{}^{s}\,\boldsymbol{v}_{\boldsymbol{i}}$$

$$\boldsymbol{x}(s) \sim {\lambda_*}^s c_i\,\boldsymbol{v}_*$$

$$s \to \infty$$

encontrar ***x*** autoconsistente se reduce a encontrar el autovector de ***A*** asociado al mayor autovalor

$$\boxed{\boldsymbol{A}\boldsymbol{x} = \lambda_*\,\boldsymbol{x}}$$

Centralidad de **autovalor**

Nodos importantes tienen **vecinos** importantes

- x_i (importancia de nodo-i) depende de x_j (importancia de nodo-j vecino). La importancia se *transmite* por la red.
- la manera más simple es asumir linealidad

$$x_i{}' = \sum_{j=1}^{N} a_{ij} x_j \qquad \boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x}$$

- el nivel de importancia se establece recursivamente en la red

$$\boldsymbol{x}'' = \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{A}\boldsymbol{x}$$

$$\vdots$$

Despues de ***s*** iteraciones: $$\boldsymbol{x}(s) = \overbrace{\boldsymbol{A}\boldsymbol{A}\ldots.\boldsymbol{A}}^{\boldsymbol{s}}\,\boldsymbol{x}(0)$$

$$\boldsymbol{x}(s) = \boldsymbol{A}^{\boldsymbol{s}}\boldsymbol{x}(0)$$

Centralidad de **autovalor**

$$\boldsymbol{Ax}^* = \lambda_* \, \boldsymbol{x}^*$$

x se conoce como **centralidad de Bonacich**

$$x_i^{\,*} = \frac{1}{\lambda_*} \sum_i A_{ij} \, x_j^{\,*}$$

La centralidad de un nodo depende de la de sus vecinos.

Es ***x**** una buena centralidad? Quien me asegura que sus componentes sean todas positivas, por ejemplo?

<u>Teorema de Perron-Frobenius</u>

Sea ***A*** una matriz de N x N con elementos no negativos. Entonces:

- ***A*** tiene un autovalor real dominante: $\lambda_1 > |\lambda_i| \text{ para } i = 2, \ldots, N$
- λ_1 corresponde a un autovalor para el que se puede elegir un autovector cuyas **componentes son no-negativas**
- Además vale que λ_1 está acotado por la mínima y máxima suma por columna de ***A***

$$min_j \left(\sum_{i=1}^{N} a_{ij} \right) \leq \lambda_1 \leq max_j \left(\sum_{i=1}^{N} a_{ij} \right)$$

Centralidad de **autovalor**

$$\boldsymbol{A}\boldsymbol{x}^{*} = \lambda_{*}\,\boldsymbol{x}^{*}$$

x se conoce como **centralidad de Bonacich**

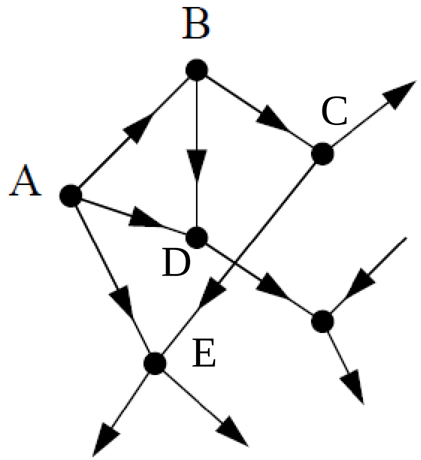
$$x_i^{\ *} = \frac{1}{\lambda_*}\sum_i A_{ij}\,x_j^{\ *}$$

La centralidad de un nodo depende de la de sus vecinos.

Para redes dirigidas (***A*** matriz no simétrica) se usa la misma definición de centralidad. Esto equivale a considerar el autovector-por-derecha de mayor autovalor.

Sin embargo puede haber problemas:

- x_A no tiene vecinos, entonces $x_A=0$
- pero también $x_B=0$, $x_C=0$, $x_D=0$, $x_E=0$

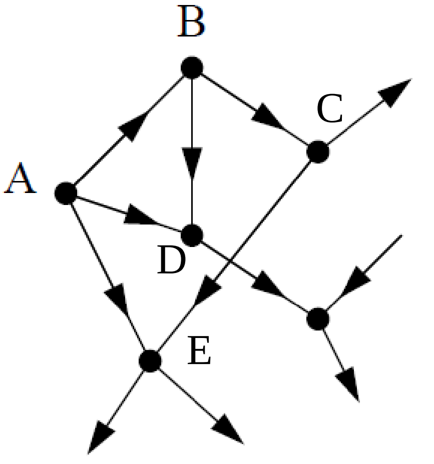


Sólo nodos de una **componente fuertemente conexa** de dos o más vértices (o la **out-component** de dicha componente) pueden tener centralidad de Bonacich no nula. Potencial problema (!)

Centralidad de **Katz**

centralidad de base.
Gran diferencia para nodos con k_{in}=0

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$



Matricialmente

$$\boldsymbol{x} = \alpha \boldsymbol{A} \boldsymbol{x} + \beta \mathbf{1}$$

$$\boldsymbol{x} - \alpha \boldsymbol{A} \boldsymbol{x} = \beta \mathbf{1}$$

$$(\boldsymbol{I} - \alpha \boldsymbol{A}) \boldsymbol{x} = \beta \mathbf{1}$$

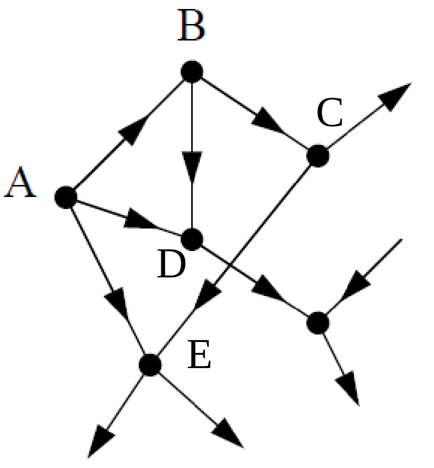
$$\boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \beta \mathbf{1}$$

Usualmente se toma β=1 y se debe especificar α, que controla la importancia relativa del primer término respecto del segundo

Centralidad de **Katz**

$$x_i = \alpha \sum_j A_{ij} x_j + 1$$

$$\boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \mathbf{1}$$



Se debe especificar α , como lo elijo?

- α =0 resulta en distribución uniforme de centralidad
- α>0 le doy más peso al primer término
- **PERO OJO:** para algún α > 0 la matriz ***I***- α ***A*** puede no ser inversible.
  Esto sucede cuando $det(\boldsymbol{I} - \alpha \boldsymbol{A}) = 0$, o lo que es lo mismo, cuando:

$$det(\boldsymbol{A} - \alpha^{-1} \boldsymbol{I}) = 0$$

La condición para identificar α problematicos asociada a la ecuación de autovalores para ***A*** (!)

Para α creciente el *primer* problema lo encuentro cuando $\alpha^{-1} = \lambda_*$

Por lo tanto, se suele elegir α en el rango $0 < \alpha < 1/\lambda_*$

Centralidad de **Katz**

Estimación de centralidad

	Algebraica:	Iterando:	
	$$\boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \beta \boldsymbol{1}$$	$$x_i^{(t+1)} = \alpha \sum_j A_{ij} x_j^{(t)} + \beta$$	
Costo computacional	o(n^3)	o(M.r)	M:#enlaces, i.e. items no nulos de ***A*** r: #iteraciones

Centralidad de **Katz generalizada**

Asignamos una relevancia-a-priori, que puede ser diferente para cada nodo, utilizando **información externa a la red**

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)$$

$$x_i^{(t+1)} = \alpha \sum_j A_{ij} x_j^{(t)} + \beta_i \qquad \boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \boldsymbol{\beta}$$

Retomando…

- Centralidad de autovalor

Nodos importantes tienen **vecinos** importantes

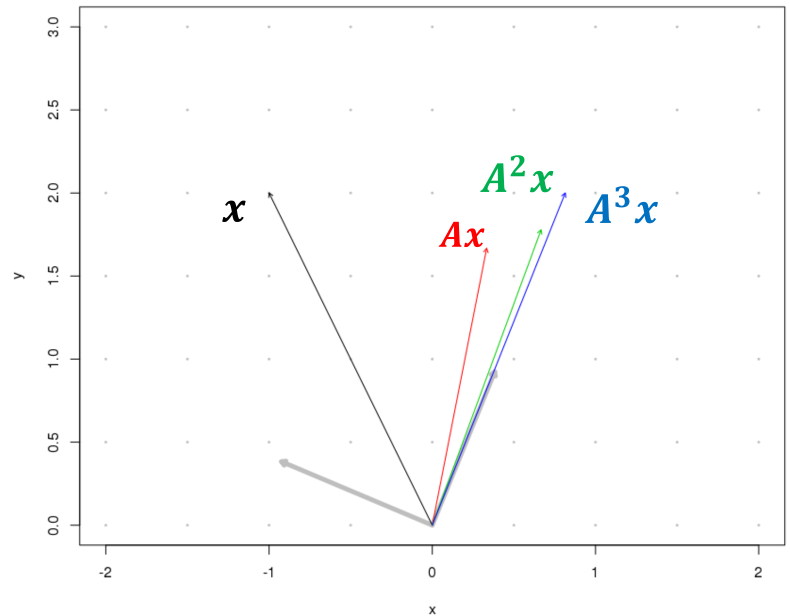
- x_i (importancia de nodo-i) depende de x_j (importancia de nodo-j vecino).
- La importancia se *transmite* por la red.
- la manera más simple es asumir linealidad

$$x_i{}' = \sum_{j=1}^{N} a_{ij} x_j \qquad \boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x}$$

$$\boldsymbol{A}\boldsymbol{x}^* = \lambda_* \, \boldsymbol{x}^*$$

Autovector del autovalor dominante de ***A***

X* se conoce como **centralidad de Bonacich**



Retomando…

- Centralidad de autovalor (Bonacich)

$$\boldsymbol{A}\boldsymbol{x}^* = \lambda_* \, \boldsymbol{x}^*$$

- Centralidad de Katz generalizada

$$x_i^{(t+1)} = \alpha \sum_j A_{ij} x_j^{(t)} + \beta_i \qquad \boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \boldsymbol{\beta}$$

$0 < \beta_i < 1$

$0 < \alpha < 1/\lambda_*$

Centralidad **PageRank**

Katz generalizado:

$$x_i^{(t+1)} = \alpha \sum_j A_{ij} x_j^{(t)} + \beta_i$$

Un posible problema: nodos con **alta centralidad** distribuyen ese valor (**alto**) a todos sus contactos k_{out}. Quizas no sea eso lo que uno quiere.

'diluye' la contribución del nodo-j

$$x_i^{(t+1)} = \alpha \sum_j \frac{1}{k_j^{out}} A_{ij} x_j^{(t)} + \beta_i$$

Centralidad **PageRank**

$$x_i{}^{(t+1)} = \alpha \sum_j \frac{1}{k_j{}^{out}} A_{ij} x_j{}^{(t)} + \beta_i$$

$$\boldsymbol{x} = \alpha \boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{x} + \boldsymbol{\beta}$$

$$\left[\boldsymbol{D}^{-1}\right]_{ii} = 1/k_i{}^{out}$$

$$\left(\boldsymbol{I} - \alpha \boldsymbol{A}\boldsymbol{D}^{-1}\right)\boldsymbol{x} = \boldsymbol{\beta}$$

$$\boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A}\boldsymbol{D}^{-1})^{-1}\boldsymbol{\beta}$$

$$= (\boldsymbol{D}\boldsymbol{D}^{-1} - \alpha \boldsymbol{A}\boldsymbol{D}^{-1})^{-1}\boldsymbol{\beta}$$

$$= \left[(\boldsymbol{D} - \alpha \boldsymbol{A})\boldsymbol{D}^{-1}\right]^{-1}\boldsymbol{\beta}$$

$$\boxed{\boldsymbol{x} = \boldsymbol{D}(\boldsymbol{D} - \alpha \boldsymbol{A})^{-1}\boldsymbol{\beta}}$$

razonando como antes $\alpha < 1/\lambda_{\text{dominante}}$ de $\boldsymbol{A}\boldsymbol{D}^{-1}$.
igual….se suele usar α=0.85

Centralidad **PageRank sin bias**

$$x_i^{(t+1)} = \cancel{\alpha} \sum_j \frac{1}{k_j^{out}} A_{ij} x_j^{(t)} + \cancel{\beta_i}$$

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{x} \qquad\qquad [\boldsymbol{D}^{-1}]_{ii} = 1/k_i^{out}$$

- Estoy buscando el autovector de $\boldsymbol{A}\boldsymbol{D}^{-1}$ asociado a *λ=1*
- Para redes no-dirigidas es fácil ver que $x_j = k_j$ es la solución autoconsistente buscada.
- O sea PageRank sin bias equivale a centralidad de grado.

Nota:
- para redes no dirigidas se puede demostrar (practica) que el autovalor dominante de AD^{-1} es 1, con autovector (k1,k2,…,kn)
- para redes dirigidas no. pero resulta o(1)

Centralidades de recurrencia (resumen)

$$x_i^{(t+1)} = \alpha \sum_j \frac{1}{k_j^{out}} A_{ij} x_j^{(t)} + \beta_i$$

$$\boldsymbol{x} = \boldsymbol{D}(\boldsymbol{D} - \alpha \boldsymbol{A})^{-\boldsymbol{1}} \boldsymbol{\beta}$$

PageRank centrality

$$x_i^{(t+1)} = \sum_j \frac{1}{k_j^{out}} A_{ij} x_j^{(t)}$$

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{D}^{-\boldsymbol{1}} \boldsymbol{x}$$

Degree centrality

$$x_i^{(t+1)} = \alpha \sum_j A_{ij} x_j^{(t)} + \beta_i$$

$$\boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-\boldsymbol{1}} \boldsymbol{\beta}$$

Katz centrality

$$x_i^{(t+1)} = \sum_j A_{ij} x_j^{(t)}$$

$$\boldsymbol{x} = \lambda_*^{-\boldsymbol{1}} \boldsymbol{A}\boldsymbol{x}$$

Eigenvector (Bonacich) centrality

Conectores (*hubs*) y Autoridades

- Hasta ahora, en **redes dirigidas**, los algoritmos que vimos asignaban más nivel de **centralidad** a aquellos nodos que **recibían** muchas conecciones

- Puede ser de interés, sin embargo, identificar nodos que actúen como buenos *reviews.* Implica idea de centralidad asociada con k_{out} hacia nodos relevantes (e.g. página con links hacia otras páginas relevantes)

Idea [Kleinberg, Authoritative Sources in a Hyperlinked Environment, 1999 10122 citas)]: Generalizar la centralidad de Bonacich y permitir que cada nodo tenga dos atributos

1. **Autoridad**: que tanto conocimiento, información, etc, tiene un nodo respecto a un tema
2. **Conectividad (*hubiness*)**: que tanto un nodo es capaz de encontrar información sobre un tema

En este esquema:

- Los nodos son caracterizados <u>simultaneamente</u> como **conectores** y **autoridades**
- Los mejores conectores apuntan a las mejores autoridades
- Algoritmo HITS (*Hyperlink-induced topic search*)

Conectores (*hubs*) y Autoridades

Para un nodo, dos centralidades: x_i (autoridad), y_i (conector)

$$x_i = \alpha \sum_j A_{ij} y_j \qquad y_i = \beta \sum_j A_{ji} x_j$$

$$\boldsymbol{x} = \alpha \boldsymbol{A} \boldsymbol{y} \qquad \boldsymbol{y} = \beta \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}$$

pero entonces

$$\boldsymbol{x} = (\alpha\beta) \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x} \qquad \boldsymbol{y} = (\alpha\beta) \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}\, \boldsymbol{y}$$

$$\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x} = \lambda\, \boldsymbol{x} \qquad \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A} \boldsymbol{y} = \lambda\, \boldsymbol{y}$$

con $\lambda = \alpha\beta$

fijarse ademas que:

- $\boldsymbol{AA^T}$ es la matriz de co-citas, por lo que la **centralidad de autoridad** puede verse como **centralidad de autovector de la red de co-citas**

- $\boldsymbol{A^TA}$ es la matriz bibliográfica, por lo que la **centralidad de conectividad** puede verse como **centralidad de autovector de la red bibliográfica**

Conectores (*hubs*) y Autoridades

Para un nodo, dos centralidades: x_i (autoridad), y_i (conector)

$$x_i = \alpha \sum_j A_{ij} y_j \qquad\qquad y_i = \beta \sum_j A_{ji} x_j$$

$$\boldsymbol{x} = \alpha \boldsymbol{A} \boldsymbol{y} \qquad\qquad \boldsymbol{y} = \beta \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}$$

pero entonces

$$\boldsymbol{x} = (\alpha\beta) \boldsymbol{A} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x} \qquad\qquad \boldsymbol{y} = (\alpha\beta) \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}\, \boldsymbol{y}$$

$$\boldsymbol{A} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x} = \lambda\, \boldsymbol{x} \qquad\qquad \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A} \boldsymbol{y} = \lambda\, \boldsymbol{y}$$

con $\lambda = \alpha\beta$

fijarse ademas que:

$$\boldsymbol{A}^{\boldsymbol{T}} (\boldsymbol{A} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}) = \lambda \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}$$

$$\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A} \left(\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}\right) = \lambda (\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x})$$

por lo que $\boldsymbol{y} = (\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x})$ o sea, basta con resolver la ecuacion de autovalores de $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$

Conectores (*hubs*) y Autoridades

Para un nodo, dos centralidades: x_i (autoridad), y_i (conector)

$$x_i = \alpha \sum_j A_{ij} y_j \qquad\qquad y_i = \beta \sum_j A_{ji} x_j$$

$$\boldsymbol{x} = \alpha \boldsymbol{A} \boldsymbol{y} \qquad\qquad \boldsymbol{y} = \beta \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}$$

pero entonces

$$\boldsymbol{x} = (\alpha\beta) \boldsymbol{A} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x} \qquad\qquad \boldsymbol{y} = (\alpha\beta) \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}\, \boldsymbol{y}$$

$$\boldsymbol{A} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x} = \lambda\, \boldsymbol{x} \qquad\qquad \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A} \boldsymbol{y} = \lambda\, \boldsymbol{y}$$

con $\lambda = \alpha\beta$

- ***AA**^{T}* es una matriz simétrica, semi-definida positiva, por lo que sus autovalores son mayores que cero.
- De hecho, sus autovalores son los valores-singulares de ***A*** elevados al cuadrado
- Perron-Frobenius: el autovector del autovalor dominante tiene componentes no-negativas