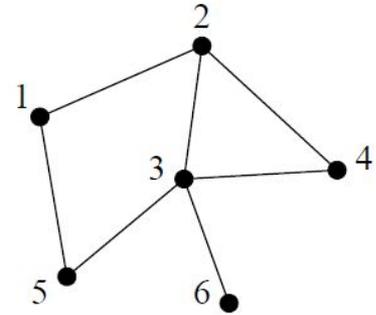


Caminatas al azar



- Imaginemos un caminante caminando por una red (simple y conexa / fuertemente conexa) de manera **aleatoria**.
 - Consideremos el tiempo de manera discreta.
 - A $t=0$ se encuentra en el nodo- k
 - Al pasar un intervalo Δt Juan avanza desde el nodo donde se encuentra actualmente hacia alguno de sus vecinos



Cual es la probabilidad, $p_i(t)$, que a tiempo t el caminantes se encuentre en el nodo- i ?

- Inicialmente Juan está en el nodo- s . El **campo** de probabilidades $\mathbf{p}(t=0)$ es

$$p_i^{(t=0)} = \delta_{si} = \begin{cases} 1 & \text{si } i = s \\ 0 & \text{si } i \neq s \end{cases}$$

- Por cómo es el proceso de desplazamiento $\mathbf{p}(t+1)$ va a ser función de $\mathbf{p}(t)$

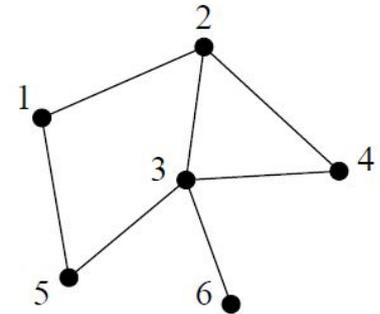
$$p_i^{(t+1)} = \sum_{j=1}^N \frac{1}{k_j} a_{ij} p_j^{(t)}$$

prob que esté en *nodo-j* a tiempo t
 prob de elegir el enlace que lleva de j a i

Donde está Juan?



$$p_i^{(t+1)} = \sum_{j=1}^N \frac{1}{k_j} a_{ij} p_j^{(t)}$$



$$\mathbf{p}^{(t+1)} = \mathbf{A} \mathbf{D}^{-1} \mathbf{p}^{(t)}$$

matriz de adyacencia con columnas normalizadas por grado del nodo

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A} \mathbf{D}^{-1} = \begin{bmatrix} 0 & 1/3 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/4 & 1/2 & 0 & 0 \\ 0 & 1/3 & 0 & 1/2 & 1/2 & 1 \\ 0 & 1/3 & 1/4 & 0 & 0 & 0 \\ 1/2 & 0 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 0 \end{bmatrix}$$

Donde está Juan?

$$p_i^{(t+1)} = \sum_{j=1}^N \frac{1}{k_j} a_{ij} p_j^{(t)}$$

$$\mathbf{p}^{(t+1)} = A D^{-1} \mathbf{p}^{(t)} \quad (1)$$

Notar que si sucede que la distribución de probabilidad converge, es decir $\mathbf{p}(t \rightarrow \infty) \rightarrow \mathbf{p}_\infty$, se debe satisfacer que

$$\mathbf{p}^{(\infty)} = A D^{-1} \mathbf{p}^{(\infty)}$$

Esto significa que \mathbf{p}_∞ es **autovector** de $A D^{-1}$, con **autovalor** 1.

Nota algebra: $\lambda = 1$ es el autovalor dominante de $A D^{-1}$

(la ecuacion (1) puede ser resuelta recursivamente)

a tiempos largos Juan está en...

$$\mathbf{p}^{(\infty)} = A D^{-1} \mathbf{p}^{(\infty)}$$

Por inspección se puede ver que el vector de probabilidades de componente i -ésima:

$$\mathbf{p}^{(\infty)}_i = \frac{k_i}{\sum_{j=1}^N k_j}$$

$$\frac{k_i}{\sum_{j=1}^N k_j} = \sum_{j=1}^N \frac{1}{k_j} a_{ij} \frac{k_j}{\sum_{r=1}^N k_r}$$

$$= \frac{1}{\sum_{r=1}^N k_r} \sum_{j=1}^N \frac{1}{k_j} a_{ij} k_j = \frac{1}{\sum_{r=1}^N k_r} \sum_{j=1}^N a_{ij}$$

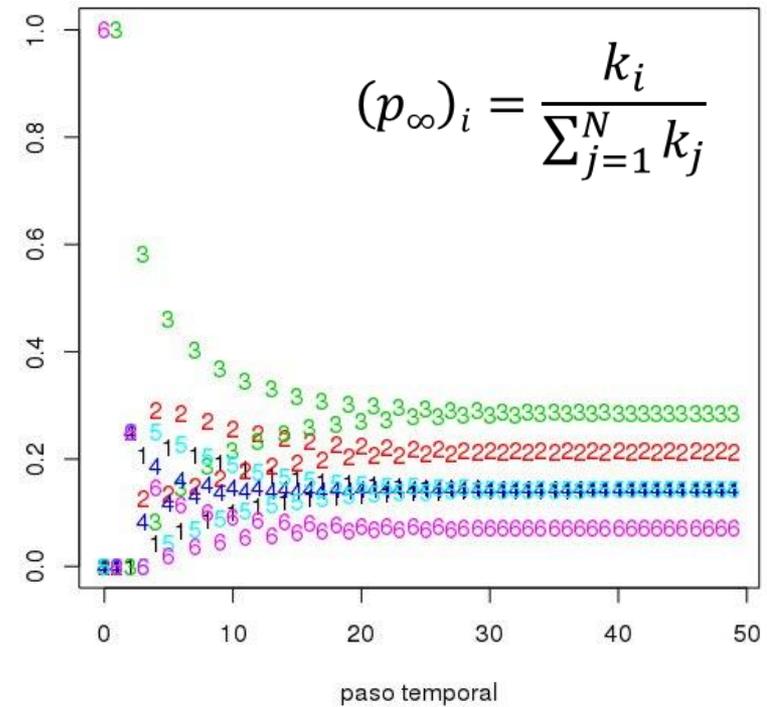
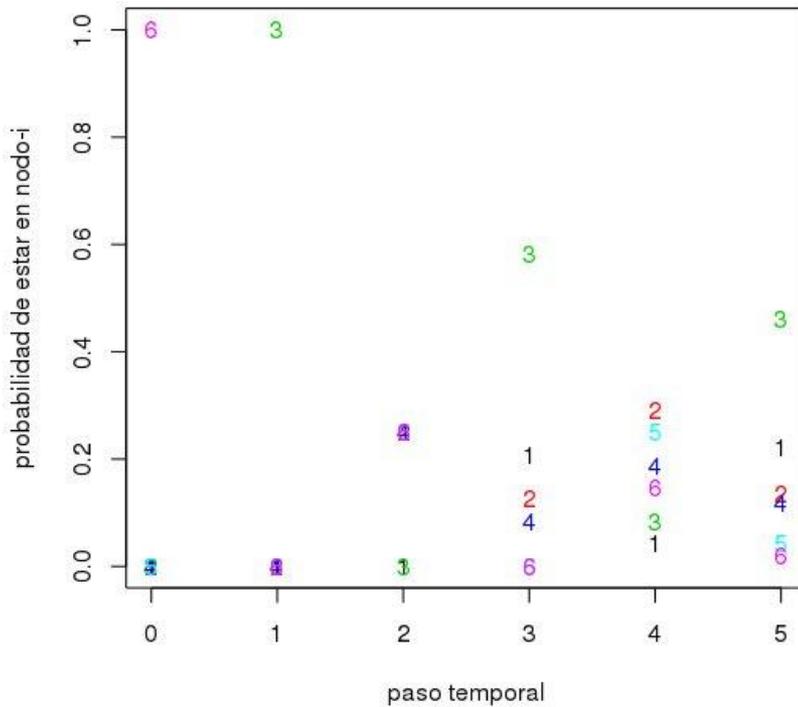
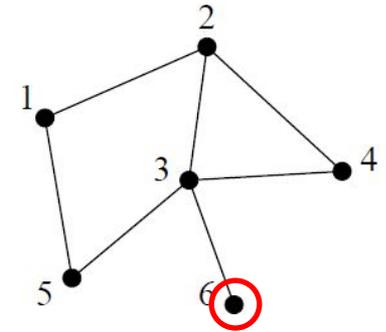
$$= \frac{k_i}{\sum_{j=1}^N k_j} \quad \checkmark$$

a tiempos largos Juan está en...

$$p_i(t+1) = \sum_{j=1}^N \frac{1}{k_j} a_{ij} p_j(t)$$

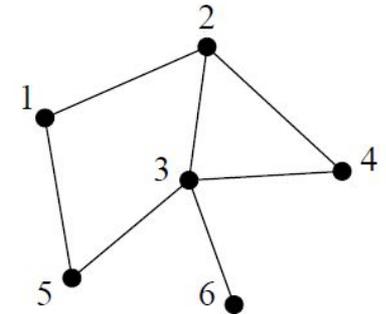
$$\mathbf{p}(0) = (0,0,0,0,0,1)^T$$

$$\mathbf{p}(t+1) = A D^{-1} \mathbf{p}(t)$$



a tiempos largos Juan está en...

$$(p_{\infty})_i = \frac{1}{\sum_{j=1}^N k_j} k_i$$



- La probabilidad de encontrar a Juan en el nodo- i es **proporcional al grado del nodo**
- En este proceso difusivo, o de caminata, las cosas tienden a **uniformizarse a nivel de enlaces**: En el régimen asintótico, la probabilidad de encontrar a Juan **atravesando uno de los M enlaces** de la red, $p_{\text{enlace}}(e_r=(i,j))$, es **uniforme**

$$(p_{\text{enlace}})_{ij} = \frac{k_j}{\sum_{r=1}^N k_r} \frac{1}{k_j} = \frac{1}{\sum_{r=1}^N k_r} = cte$$

prob de estar en j $\frac{k_j}{\sum_{r=1}^N k_r}$ prob de elegir uno de los vecinos de j $\frac{1}{k_j}$

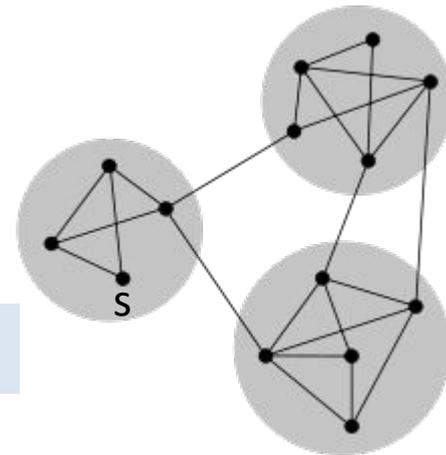
Caminatas (cortas) y similaridad

- Con qué probabilidad, en una **caminata corta**, voy a visitar un nodo-x si a tiempo $t=0$ parto de un **nodo s**?

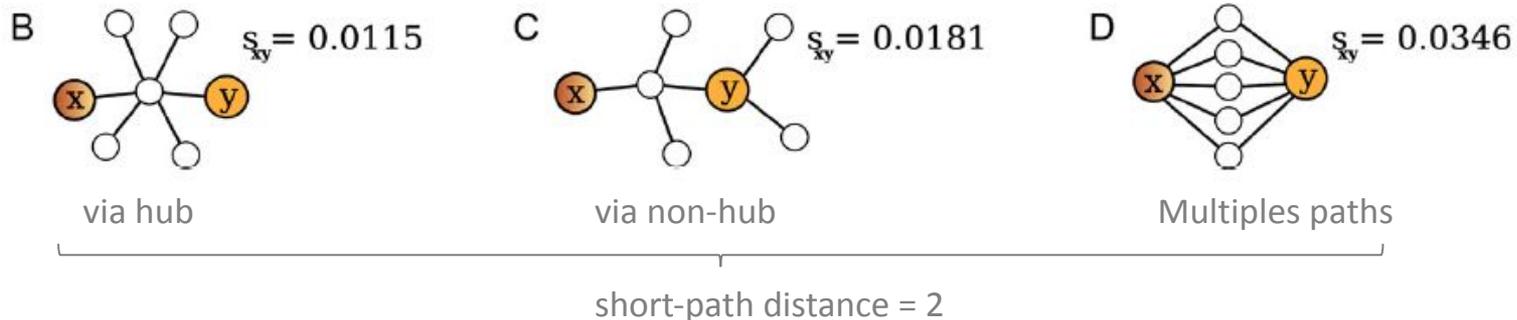
$$\mathbf{p}_s^{(t=0)} = (0, \dots, 0, 1, 0, \dots, 0)^T$$

$1, \dots, s-1, s, s+1, \dots, N$

$$\mathbf{p}_s^{(t+1)} = A D^{-1} \mathbf{p}_s^{(t)} \longrightarrow \mathbf{p}_s^{(\tau)}, \text{ con } 0 < \tau \ll \infty$$

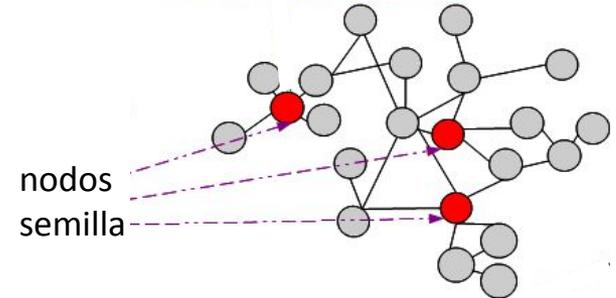


El campo $\mathbf{p}_s^{(\tau)}$ es una **medida de similaridad** de todos los nodos respecto al nodo-s



Priorización basada en redes

- Con qué probabilidad, en una **caminata corta**, voy a visitar un nodo-x si a tiempo $t=0$ parto de un **conjunto nodos de interés (nodos semilla)**?
- Con qué probabilidad **asintótica** voy a visitar un nodo-x si a tiempo $t=0$ parto de un nodo tomado al azar de un conjunto de nodos de interés (nodos semilla) y eventualmente **fuerzo revisitar semillas**?



Probabilidad de visita al nodo-x es una **medida de asociación** entre el nodo-x y el conjunto semilla

- Integro** dos espacios de conocimiento
1. el embebido en el conexionado de la red
 2. el que utilicé para definir el conjunto de nodos-semilla

Priorización de genes asociados a enfermedades

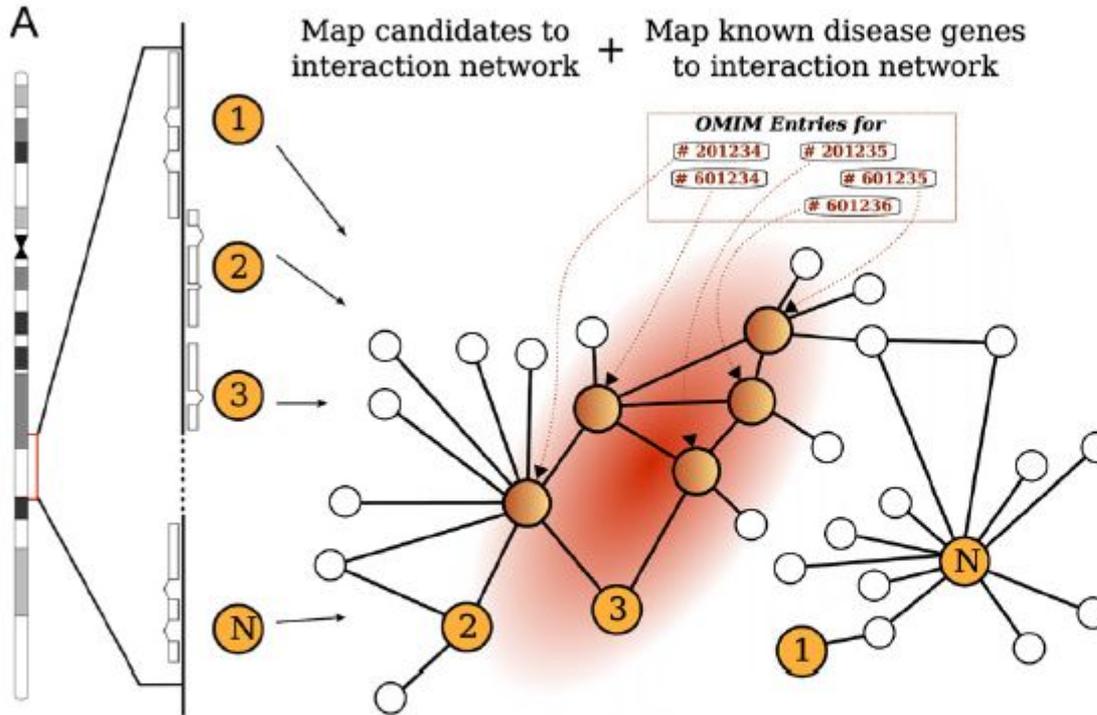
ARTICLE

Walking the Interactome for Prioritization of Candidate Disease Genes

Sebastian Köhler,^{1,2} Sebastian Bauer,^{1,2} Denise Horn,¹ and Peter N. Robinson^{1,*}

The identification of genes associated with hereditary disorders has contributed to improving medical care and to a better understanding of gene functions, interactions, and pathways. However, there are well over 1500 Mendelian disorders whose molecular basis remains unknown. At present, methods such as linkage analysis can identify the chromosomal region in which unknown disease genes are located, but the regions could contain up to hundreds of candidate genes. In this work, we present a method for prioritization of candidate genes by use of a global network distance measure, random walk analysis, for definition of similarities in protein-protein interaction networks. We tested our method on 110 disease-gene families with a total of 783 genes and achieved an area under the ROC curve of up to 98% on simulated linkage intervals of 100 genes surrounding the disease gene, significantly outperforming previous methods based on local distance measures. Our results not only provide an improved tool for positional-cloning projects but also add weight to the assumption that phenotypically similar diseases are associated with disturbances of subnetworks within the larger protein interactome that extend beyond the disease proteins themselves.

Priorización de nuevas asociaciones gen/enfermedad



Algoritmos para **propagar sentido de pertenencia** al conjunto de interés:

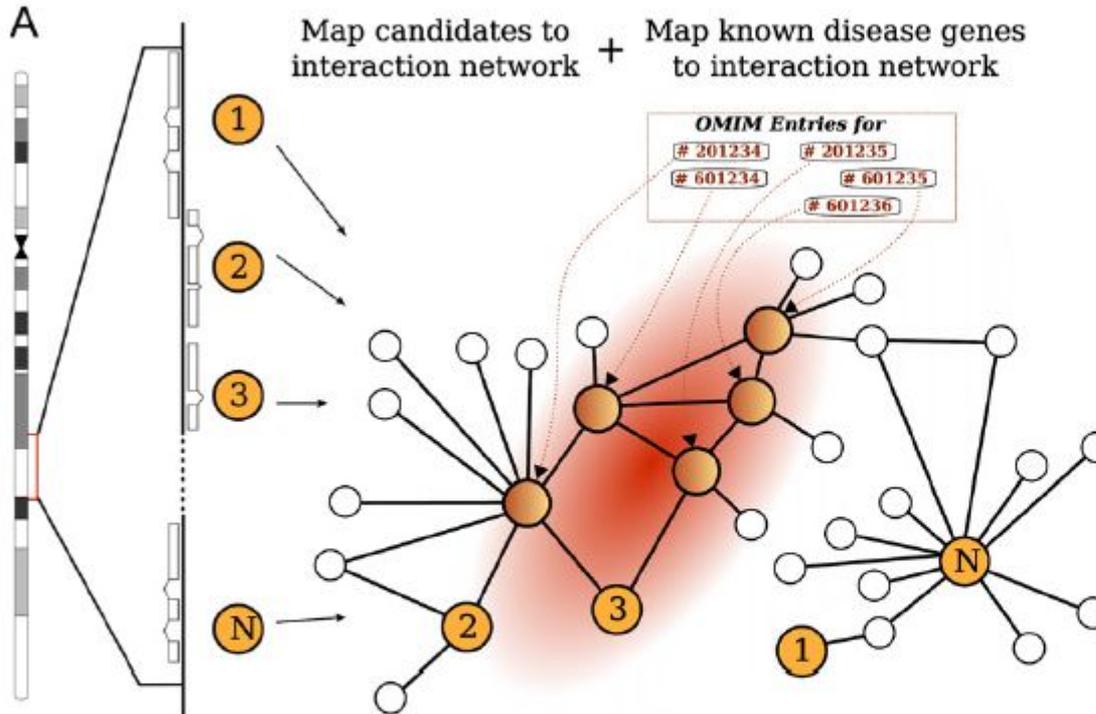
Random Walk with Restart (Kholer 2009)

Net-Rank (Chen 2009)

Net-Propagation (Vanunu 2010)

Functional Flow (Navieba 2005)

Priorización de nuevas asociaciones gen/enfermedad



Algoritmos para **propagar sentido de pertenencia** al conjunto de interés:

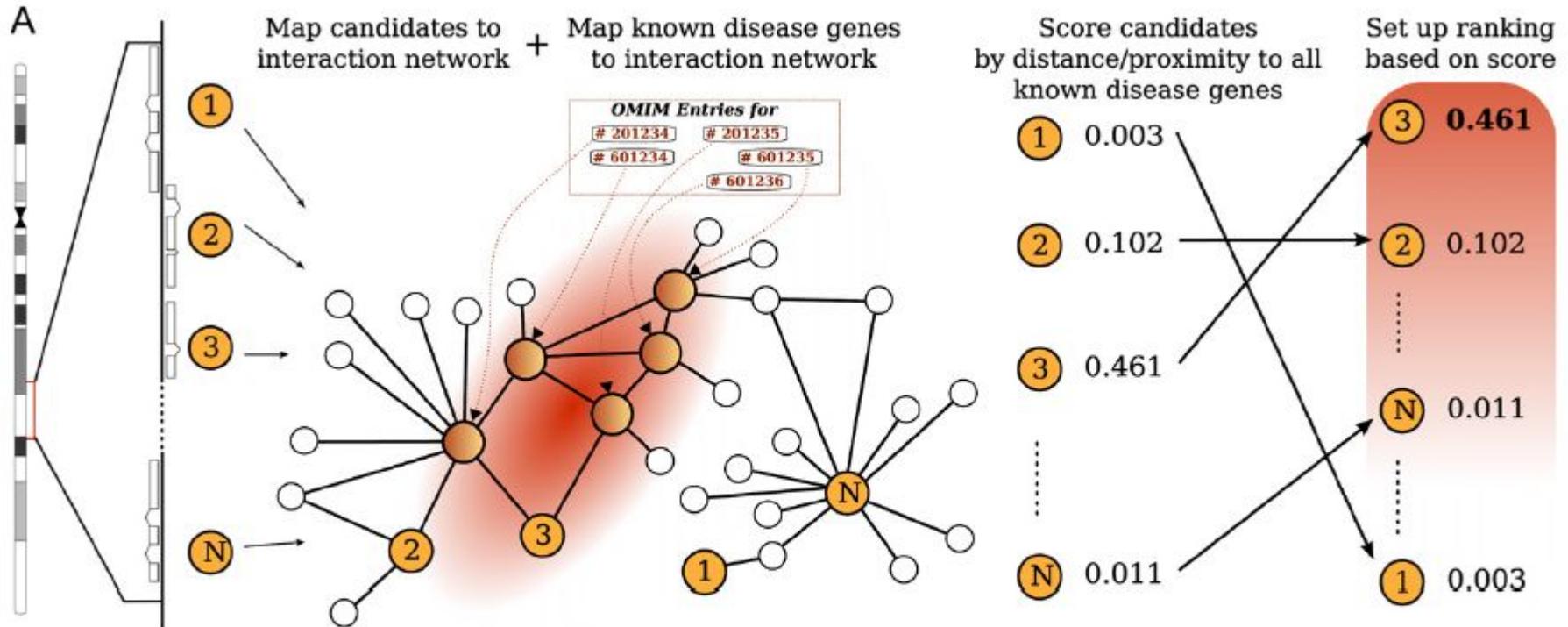
Random Walk with Restart (Kholer 2009)

$$p_i^{t+1} = (1 - r) \sum_{j=1}^N a_{ij} p_j^t + r p_i^{t=0}$$

$$\mathbf{p}^{t+1} = (1 - r) \mathbf{A} \mathbf{p}^t + r \mathbf{p}^0$$

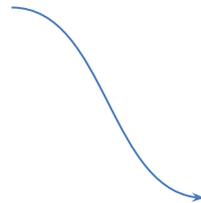
- r controla el sesgo hacia las condiciones iniciales
- el score se suele estimar a partir de simular un número reducido de pasos k

Priorización de nuevas asociaciones gen/enfermedad



Genómica funcional

- En la era post-genómica conocemos en alto grado la composición exacta del genoma de un gran número de organismos
- Sin embargo, entender qué función cumple la proteína asociada a un dado gen sigue siendo un reto
- En general los métodos computacionales utilizaban exhaustivamente el criterio de **similaridad de secuencia** para asignar funciones putativas a una dada proteína aun no caracterizada funcionalmente
- La disponibilidad de datos omicos a escala global permite abordar esto desde otra perspectiva:



**Whole-proteome prediction of protein function
via graph-theoretic analysis of interaction maps**

Elena Nabieva^{1,2}, Kam Jim², Amit Agarwal¹, Bernard Chazelle¹
and Mona Singh^{1,2,*}

¹Computer Science Department and ²Lewis-Sigler Institute for Integrative Genomics,
Princeton University, Princeton, NJ 08544, USA

Received on January 15, 2005; accepted on March 27, 2005



**Whole-proteome prediction of protein function
via graph-theoretic analysis of interaction maps**

Elena Nabieva^{1,2}, Kam Jim², Amit Agarwal¹, Bernard Chazelle¹
and Mona Singh^{1,2,*}

¹Computer Science Department and ²Lewis-Sigler Institute for Integrative Genomics,
Princeton University, Princeton, NJ 08544, USA

Received on January 15, 2005; accepted on March 27, 2005

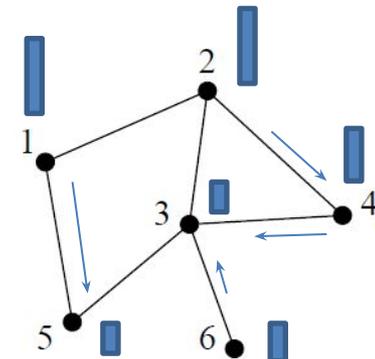
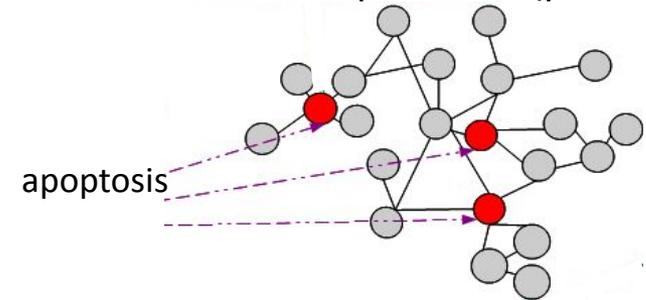
Functional Flow

Asignación de funcionalidad a proteínas aun no caracterizadas simulando flujos de información en redes de interacción de proteínas.

Dime con quien andas y te diré que haces

- Consideramos una función/proceso biológico (por ejemplo: GO-apoptosis)
- Cada **proteína anotada** a esa categoría es considerada una **fuentes de flujo** de una cantidad extensiva $x_i(t)$ que representa un *score-funcional*
- Pensamos en un proceso difusivo en el que el flujo desde el *nodo-j* al *nodo-i* es **proporcional** a la diferencia de material: $flujo_{j \rightarrow i} \sim C * (x_j - x_i)$
- Luego de cada paso, se reinyecta x a las semillas.
- La cantidad de flujo que recibe una proteína luego de k pasos de difusión representa el grado de asociación de la misma con la categoría funcional analizada.

Red de interacción de proteínas (pesada)

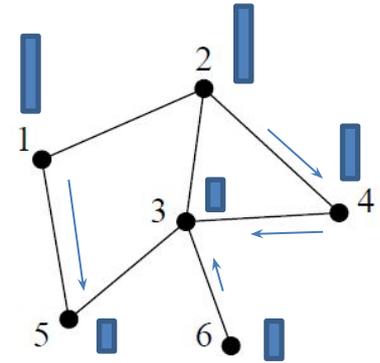


$$\frac{dx_i}{dt} = C \sum_{j=1}^N A_{ij} (x_j - x_i) \theta((x_j - x_i))$$

Difusión en redes

Pensamos en un **proceso difusivo** en el que el flujo desde el *nodo-j* al *nodo-i* es **proporcional** a la diferencia de material: $flujo_{j \rightarrow i} \sim C * (x_j - x_i)$

cte de difusión



$$\begin{aligned} \frac{dx_i}{dt} &= C \sum_{j=1}^N A_{ij} (x_j - x_i) \\ &= C \sum_{j=1}^N A_{ij} x_j - C \underbrace{\sum_{j=1}^N A_{ij} x_i}_{k_i} \\ &= C \sum_{j=1}^N A_{ij} x_j - C k_i x_i \\ &= C \sum_{j=1}^N A_{ij} x_j - C \delta_{ij} k_j x_j \end{aligned}$$

$$\frac{dx_i}{dt} = -C \sum_{j=1}^N \underbrace{(\delta_{ij} k_j - A_{ij})}_{[L]_{ij}} x_j$$

$$\frac{dx}{dt} = -C \cdot L x$$

$$\frac{dx}{dt} = C \nabla^2 x$$

Ec de difusion de calor

Referencias

- *Networks, An Introduction*. Mark Newman
- Charlas
 - Peter Dodds – Random Walk and diffusion
 - Vito Latora

Random Walk with Restart (Kholer 2009)

Net-Rank (Chen 2009)

Net-Propagation (Vanunu 2010)

Functional Flow (Navieba 2005)

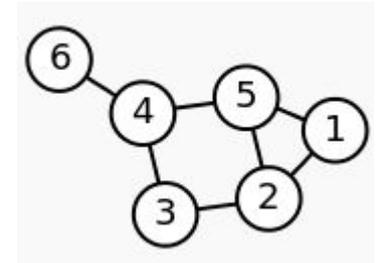
Prioritization (Lu 2012 – Physics Report)

Laplacianos y aprendizaje semisupervisado

Difusión en redes

Pensamos en un proceso difusivo en el que el flujo desde el *nodo-j* al *nodo-i* es **proporcional** a la diferencia de material: $flujo_{j \rightarrow i} \sim C * (x_j - x_i)$

└── cte de difusión



$$\frac{dx_i}{dt} = -C \sum_{j=1}^N \underbrace{(\delta_{ij}k_j - A_{ij})}_{[L]_{ij}} x_j$$

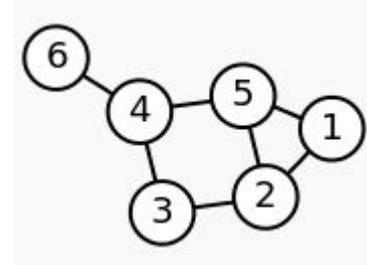
$$L = D - A$$

Degree matrix	Adjacency matrix	Laplacian matrix
$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

Propiedades

$$L = D - A$$

- L es simétrica
- L es semidefinida-positiva (i.e. autovalores $\lambda_i \geq 0, i = 1 \dots N$)
- Suma por columnas y por filas de L es cero
- Siempre $\lambda = 0$ es autovalor de L con autovector $\mathbf{v}_0 = (1, 1, \dots, 1)$, ya que $L \mathbf{v}_0 = \mathbf{0}$
- L es una matriz singular (i.e. no inversible)



Laplacian matrix

$$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

Propiedades

$$L = D - A$$

Laplaciano **combinatorio**

- L es simétrica
- L es semidefinida-positiva (i.e. autovalores $\lambda_i \geq 0, i = 1 \dots N$)
- Suma por columnas y por filas de L es cero
- Siempre $\lambda = 0$ es autovalor de L con autovector $\mathbf{v}_0 = (1, 1, \dots, 1)$, ya que $L \mathbf{v}_0 = 0 \mathbf{v}_0 = 0$
- L es una matriz singular (i.e. no inversible)
- Si el grafo tiene p componentes, de tamaños n_1, n_2, \dots, n_p

$$L = \begin{pmatrix} \boxed{L_1} & 0 & \dots \\ 0 & \boxed{L_2} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Notar que ahora habrá p autovectores de L asociados a $\lambda_{i=1, \dots, p} = 0$ de la forma

$$(\underbrace{1, \dots, 1}_{n_1}, 0, 0, \dots, 0)$$

$$(0, \dots, 0, \underbrace{1, \dots, 1}_{n_2}, 0, 0, \dots, 0)$$

$$(0, \dots, 0, 0, \dots, 0, \underbrace{1, \dots, 1}_{n_3}, 0, 0, \dots, 0)$$

Propiedades

$$L = D - A$$

Laplaciano **combinatorio**

- L es simétrica
- L es semidefinida-positiva (i.e. autovalores $\lambda_i \geq 0, i = 1 \dots N$)
- Suma por columnas y por filas de L es cero
- Siempre $\lambda = 0$ es autovalor de L con autovector $\mathbf{v}_0 = (1, 1, \dots, 1)$, ya que $L \mathbf{v}_0 = \mathbf{0}$
- L es una matriz singular (i.e. no inversible)
- Si el grafo tiene p componentes, de tamaños n_1, n_2, \dots, n_p , existen p autovectores de L asociados a $\lambda = 0$

Corolario:

- el segundo autovalor de L es $\lambda_2 \neq 0$ si el grafo posee una única componente
- λ_2 se denomina **conectividad algebraica** de la red

Autovectores de L



(a) a linear unweighted graph with two segments

$$L\phi_i = \lambda_i \phi_i$$



autovector i-esimo

ϕ_i define un campo escalar sobre el grafo

Nota1: autovectores de bajo índice están asociados a **campos más suaves** sobre el grafo

Nota2: cualquier campo puede ser escrito como CL de autovectores de L

Caminatas al azar...y Laplacianos

$$p_i(t + 1) = \sum_{j=1}^N \frac{1}{k_j} a_{ij} p_j(t)$$

$p_i(t)$ probabilidad de encontrar a Juan en el nodo- i , en el paso temporal t

- Esta misma ecuación aplica a procesos de tipo **difusivos** en una red, donde en cada paso temporal toda la cantidad de material que se encuentra en el nodo- i es repartida y enviada a sus vecinos

$$x_i(t + 1) = \sum_{j=1}^N \frac{1}{k_j} a_{ij} x_j(t)$$

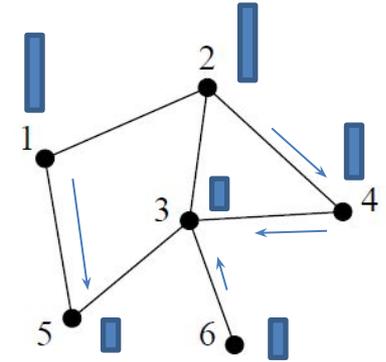
$x_j(t)$ cantidad de material que a tiempo t se encuentra en el nodo- j

el recurso x_j se reparte *extensivamente*

Difusión *random-walk* en redes

A cada paso, la cantidad x_i de cada nodo se altera

$$x_i(t + 1) = \sum_{j=1}^N \frac{1}{k_j} a_{ij} x_j(t)$$



en cada nodo hay una cantidad x_i de material

$$\Delta x_i = x_i(t + 1) - x_i(t) = \sum_{j=1}^N \frac{1}{k_j} a_{ij} x_j(t) - x_i(t)$$

$$= \sum_{j=1}^N \frac{1}{k_j} a_{ij} x_j(t) - \delta_{ij} x_j(t) \quad \delta_{ij} = 1 \text{ si } i = j$$

$$= - \sum_{j=1}^N \underbrace{\left(\delta_{ij} - \frac{1}{k_j} a_{ij} \right)}_{[L^{rw}]_{ij}} x_j(t)$$

Laplaciano *random-walk*

$$\Delta \mathbf{x} = -L^{rw} \mathbf{x}$$

$$\frac{d\mathbf{x}}{dt} = -L^{rw} \mathbf{x}$$

$$\frac{d\mathbf{x}}{dt} = \nabla^2 \mathbf{x}$$

Laplacianos

Entonces, vimos dos tipos de procesos difusivos:

$$x_i(t+1) = \sum_{j=1}^N a_{ij} x_j(t) \longrightarrow \frac{dx_i}{dt} = -C \sum_{j=1}^N \overbrace{(\delta_{ij} k_j - a_{ij})}^{[L]_{ij}} x_j \longrightarrow L = D - A$$

$$x_i(t+1) = \sum_{j=1}^N \frac{1}{k_j} a_{ij} x_j(t) \longrightarrow \frac{dx_i}{dt} = - \sum_{j=1}^N \overbrace{\left(\delta_{ij} - \frac{1}{k_j} a_{ij} \right)}^{[L^{rw}]_{ij}} x_j(t) \longrightarrow L^{rw} = I - AD^{-1} = L D^{-1}$$

$$L^{sym} \equiv D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$$

L : Laplaciano combinatorio o no-normalizado

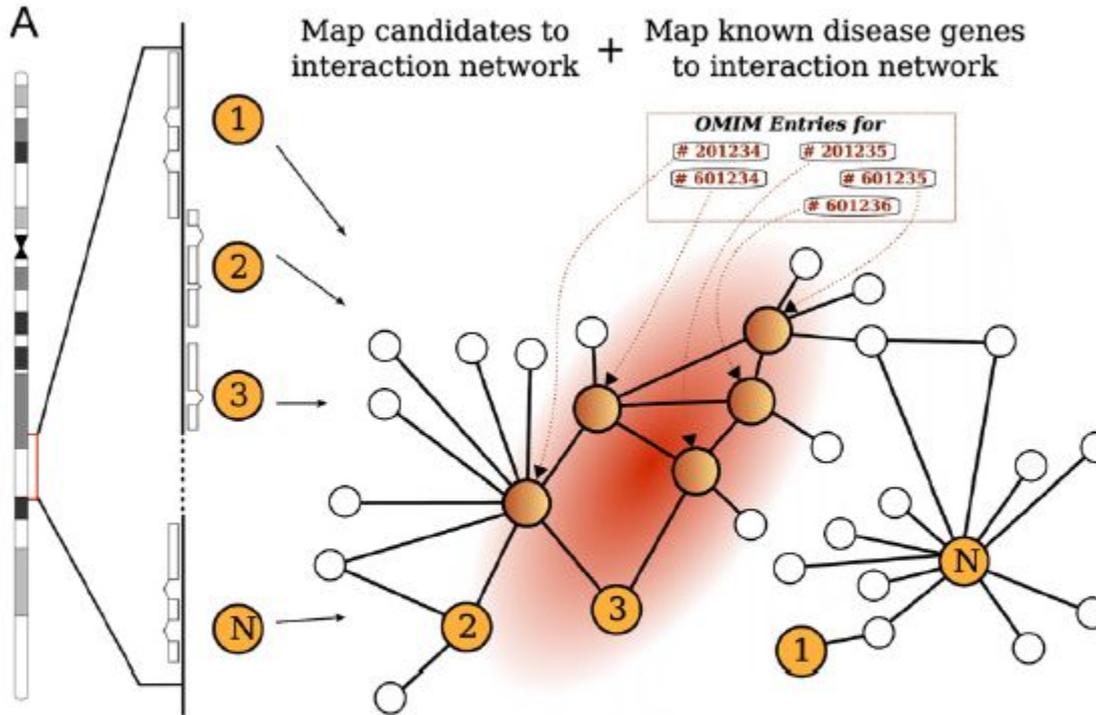
L^{rw}: Laplaciano random-walk

L^{sym}: Laplaciano normalizado o simetrico

En general

$$\dots \quad \frac{d\mathbf{x}}{dt} = -L^{[comb, rw, sym]} \mathbf{x}$$

Priorización de nuevas asociaciones gen/enfermedad



Algoritmos para **propagar sentido de pertenencia** al conjunto de interés:

Random Walk with Restart (Kholer 2009)

Net-Rank (Chen 2009)

Net-Propagation (Vanunu 2010)

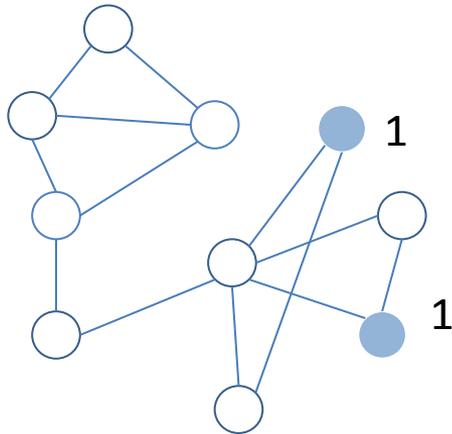
Functional Flow (Navieba 2005)

Lo podemos pensar también como un proceso de **difusión con fuentes**

$$\frac{dx}{dt} = -L^{[comb,rw,sym]}x$$

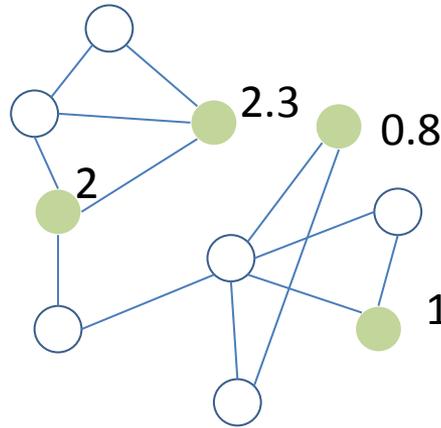
Difusión en redes y aprendizaje semi-supervisado

Tengo información parcial asociada a un subconjunto de nodos (etiquetas, o valores reales) y quiero utilizarla para inferir propiedades de los no-etiquetados. Cada nodo va a propagar su etiqueta de manera iterativa hasta converger.



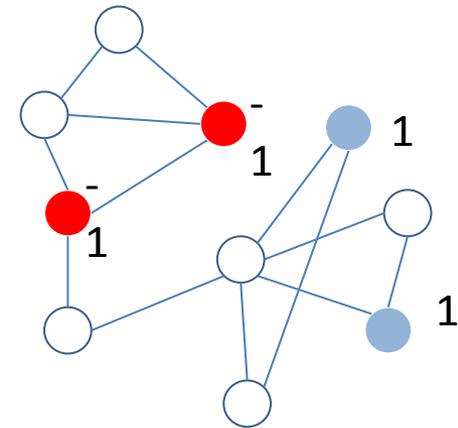
Problema de priorización:

l nodos etiquetados con valor 1 y $N-l$ nodos con valor 0.



Problema de regresión:

l nodos etiquetados con valores reales y $N-l$ nodos con valor 0.



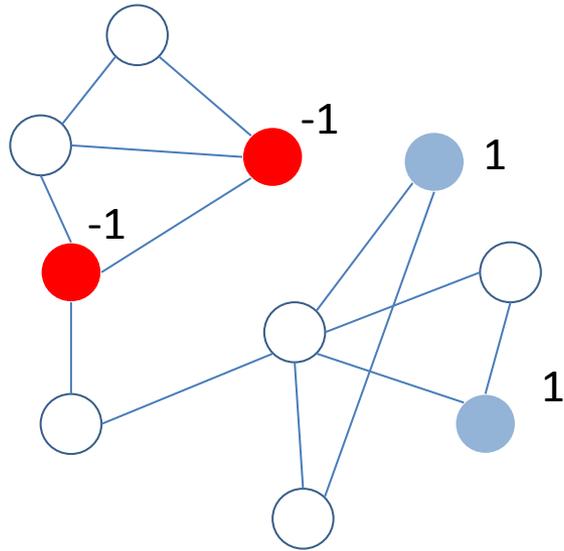
Problema de clasificación:

l nodos etiquetados con valor 1 (azul) o -1 (rojo) y $N-l$ nodos con valor 0. Nota: para multiclase se define campo vectorial sobre el grafo.

El problema de aprendizaje **semi-supervisado** consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red (smoothness or manifold assumption)

Difusión en redes y aprendizaje semi-supervisado



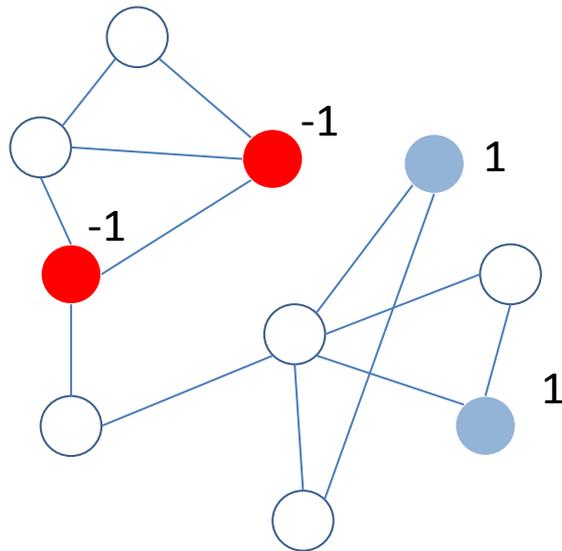
Algoritmo 1 Label Propagation (Zhu 2002)

- Computo de matriz de adyacencia W
- Computo de matriz diagonal D : $D_{ii} \leftarrow \sum_j w_{ij}$
- Inicializo $\hat{Y}^{(t=0)} \leftarrow (y_1, \dots, y_l, 0, 0, \dots, 0)$
- Itero hasta convergencia
 1. $\hat{Y}^{(t+1)} \leftarrow D^{-1}W\hat{Y}^{(t)}$
 2. $\hat{Y}_l^{(t+1)} \leftarrow Y_l$
- Etiqueta del nodo- i resulta $sign(\hat{y}_i^{(\infty)})$

Sea el grafo G ,

- nodos $1, 2, \dots, l$ etiquetados no trivialmente según $Y_l = (y_1, \dots, y_l)$
- nodos $l+1, \dots, N$ etiquetados con valor 0
- Queremos **propagar la información** por la red y estimar el vector de **etiquetas** asintótico: $\hat{Y} = (\hat{Y}_l, \hat{Y}_u)$

Difusión en redes y aprendizaje semi-supervisado



Algoritmo 2 Label Propagation

- Computo de matriz de adyacencia W , se fija $w_{ii}=0$
- Computo de matriz diagonal D : $D_{ii} \leftarrow \sum_j w_{ij}$
- Elijo $\epsilon > 0$ y $\alpha \in (0,1)$

$$\mu \leftarrow \frac{\alpha}{1-\alpha} (0, +\infty)$$
- Computo la matriz diagonal $A_{ii} \leftarrow I_{[l]}(i) + \mu D_{ii} + \mu \epsilon$
- Inicializo $\hat{Y}^{(t=0)} \leftarrow (y_1, \dots, y_l, 0, 0, \dots, 0)$
- Itero hasta convergencia

$$\hat{Y}^{(t+1)} \leftarrow A^{-1}(\mu W \hat{Y}^{(t)} + \hat{Y}^{(0)})$$
- Etiqueta del nodo- i resulta $sign(\hat{y}_i^{(\infty)})$

para nodo- i etiquetado

$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j W_{ij} \hat{y}_j^{(t)} + \frac{1}{\mu} y_i}{\sum_j W_{ij} + \frac{1}{\mu} + \epsilon}$$

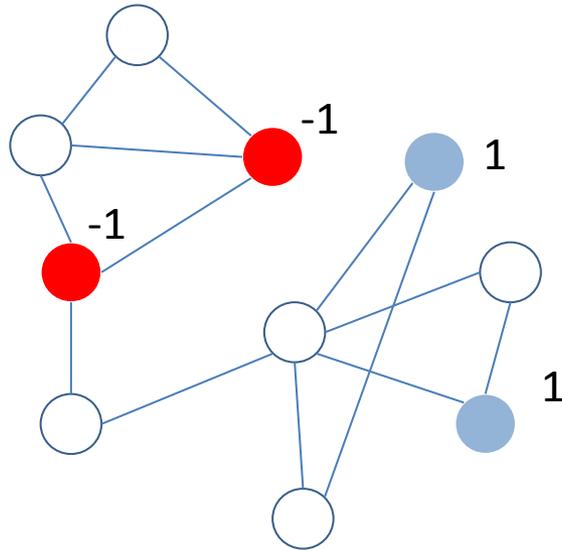
para nodo- i sin etiquetar

$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j W_{ij} \hat{y}_j^{(t)}}{\sum_j W_{ij} + \epsilon}$$

$I_{[l]}$: matriz diagonal con 1's en los primeros l elementos y 0 el resto

$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j W_{ij} \hat{y}_j^{(t)} + \frac{1}{\mu} y_i}{\left(\frac{1}{\mu} I_{[l]} + \sum_j W_{ij} + \epsilon\right)}$$

Difusión en redes y aprendizaje semi-supervisado



Algoritmo 2 Label Propagation

- Computo de matriz de adyacencia W , se fija $w_{ii}=0$
- Computo de matriz diagonal D : $D_{ii} \leftarrow \sum_j w_{ij}$
- Elijo $\epsilon > 0$ y $\alpha \in (0,1)$

$$\mu \leftarrow \frac{\alpha}{1-\alpha} (0, +\infty)$$
- Computo la matriz diagonal $A_{ii} \leftarrow I_{[l]}(i) + \mu D_{ii} + \mu \epsilon$
- Inicializo $\hat{Y}^{(t=0)} \leftarrow (y_1, \dots, y_l, 0, 0, \dots, 0)$
- Itero hasta convergencia

$$\hat{Y}^{(t+1)} \leftarrow A^{-1}(\mu W \hat{Y}^{(t)} + \hat{Y}^{(0)})$$
- Etiqueta del nodo- i resulta $sign(\hat{y}_i^{(\infty)})$

Notar:

- $\mu \rightarrow 0$ implica $\hat{Y}^{(t+1)} \leftarrow \hat{Y}^{(0)}$
- Se permite $\hat{Y}_l \neq Y_l$
- Se considera un termino de regularización ϵ

para nodo- i etiquetado

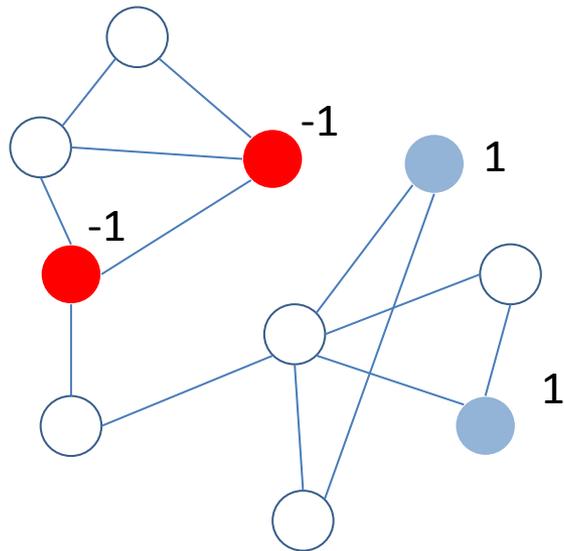
$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j W_{ij} \hat{y}_j^{(t)} + \frac{1}{\mu} y_i}{\sum_j W_{ij} + \frac{1}{\mu} + \epsilon}$$

para nodo- i sin etiquetar

$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j W_{ij} \hat{y}_j^{(t)}}{\sum_j W_{ij} + \epsilon}$$

$I_{[l]}$: matriz diagonal con 1's en los primeros l elementos y 0 el resto

Difusión en redes y aprendizaje semi-supervisado



Algoritmo 3 Label Propagation (Zhou 2004)

- Computo de matriz de adyacencia W , se fija $w_{ii}=0$
- Computo de matriz diagonal D : $D_{ii} \leftarrow \sum_j w_{ij}$
- Computo *Laplaciano simetrico (notacion)*

$$\mathcal{L} \leftarrow D^{-1/2} W D^{-1/2}$$

- Inicializo $\hat{Y}^{(t=0)} \leftarrow (y_1, \dots, y_l, 0, 0, \dots, 0)$
- Elijo $\alpha \in [0, 1)$
- Itero hasta convergencia

$$\hat{Y}^{(t+1)} \leftarrow \alpha \mathcal{L} \hat{Y}^{(t)} + (1 - \alpha) \hat{Y}^{(0)}$$

- Etiqueta del nodo- i resulta $\text{sign}(\hat{y}_i^{(\infty)})$

Buscando buenas soluciones

El problema de aprendizaje supervisado consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red

Buenas y suaves soluciones

El problema de aprendizaje supervisado consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red

I. Consistencia con etiquetado inicial:

Elementos etiquetados originalmente

Sea $\hat{Y} = (\hat{Y}_l, \hat{Y}_u)$

no-etiquetados originalmente

Notar el limite: la diferencia la estimo considerando los elementos conocidos

$$\sum_{i=1}^l (\hat{y}_i - y_i)^2 = \|\hat{Y}_l - Y_l\|^2$$

Etiquetado estimado

Etiquetado real

Buenas y suaves soluciones

El problema de aprendizaje supervisado consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red

Elementos etiquetados

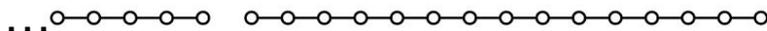
Sea $\hat{Y} = (\hat{Y}_l, \hat{Y}_u)$

↓
no-etiquetados

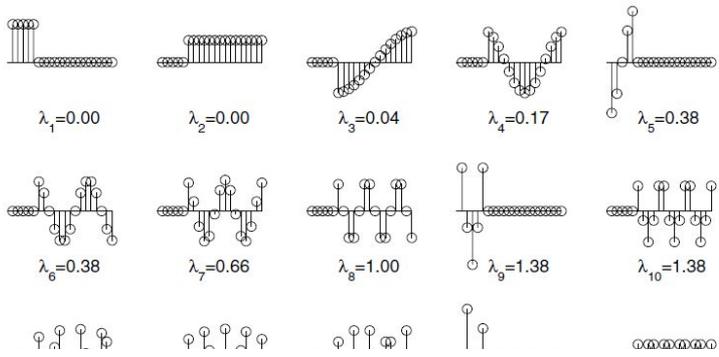
Consistencia con etiquetado inicial:

$$\sum_{i=1}^l (\hat{y}_i - y_i)^2 = \|\hat{Y}_l - Y_l\|^2$$

Laplaciano



(a) a linear unweighted graph with two segments



funcion **suave**

$$\check{Y} = \sum_{i=1}^p a_i \phi_i$$

$p < n$

Cualquier campo puede ser escrito como:

$$F = \sum_{i=1}^N a_i \phi_i \quad \text{con} \quad L\phi_i = \lambda_i \phi_i$$

$(\lambda_i < \lambda_{i+1})$

La solución buscada:

$$\check{Y}^* = \operatorname{argmin}_{\{a_1 \dots a_p\}} \|\check{Y}_l - Y_l\|^2$$

Otra manera...no tan espectral....regularización en grafos

El problema de aprendizaje supervisado consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red

etiquetados

Sea $\hat{Y} = (\hat{Y}_l, \hat{Y}_u)$

no-etiquetados

penalizamos apartamiento de etiquetado original

Consistencia con etiquetado inicial: $\sum_{i=1}^l (\hat{y}_i - y_i)^2 = \|\hat{Y}_l - Y_l\|^2$

Consistencia con estructura de conectividad de los datos:

$$\frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\hat{y}_i - \hat{y}_j)^2$$

penalizamos cambios **bruscos** de etiquetado

Otra manera...regularización en grafos

El problema de aprendizaje supervisado consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red

etiquetados

Sea $\hat{Y} = (\hat{Y}_l, \hat{Y}_u)$

no-etiquetados

penalizamos apartamiento de etiquetado original

Consistencia con etiquetado inicial:

$$\sum_{i=1}^l (\hat{y}_i - y_i)^2 = \|\hat{Y}_l - Y_l\|^2$$

Consistencia con estructura de conectividad de los datos:

$$\frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\hat{y}_i - \hat{y}_j)^2 = \frac{1}{2} \left(2 \sum_{i=1}^n \hat{y}_i^2 \sum_{j=1}^n \mathbf{W}_{ij} - 2 \sum_{i,j=1}^n \mathbf{W}_{ij} \hat{y}_i \hat{y}_j \right)$$

penalizamos cambios **bruscos** de etiquetado

$$= \hat{Y}^\top (\mathbf{D} - \mathbf{W}) \hat{Y}$$

$$= \hat{Y}^\top L \hat{Y}$$

La forma cuadratica inducida por L asegura suavidad!

Regularización en grafos

El problema de aprendizaje supervisado consiste en encontrar un etiquetado de los nodos del grafo consistente con

- I. El etiquetado inicial (incompleto)
- II. La estructura de la red

penalizamos apartamiento de etiquetado original

penalizamos cambios **bruscos** de etiquetado

$$\mathcal{H}(\hat{Y}) = \sum_{i=1}^l (\hat{y}_i - y_i)^2 + \frac{1}{2} \sum_{i,j=1}^N w_{ij} (\hat{y}_i - \hat{y}_j)^2$$

$$= \|\hat{Y}_l - Y_l\|^2 + \hat{Y}^T L \hat{Y}$$

Se trata de encontrar el etiquetado \hat{Y} que minimice $\mathcal{H}(\hat{Y})$

En la practica se suele agregar un término de regularización (que representa un costo basal de etiquetado, para romper degeneracion)

$$\mathcal{H}(\hat{Y}) = \|\hat{Y}_l - Y_l\|^2 + \mu \hat{Y}^T L \hat{Y} + \mu \epsilon \|\hat{Y}\|^2$$

Buscando buenas soluciones

$$\mathcal{H}(\hat{Y}) = \|\hat{Y}_l - Y_l\|^2 + \mu \hat{Y}^T L \hat{Y} + \mu \epsilon \|\hat{Y}\|^2 \quad (1)$$

- Estrategia 1:
 - ✓ Forzar las etiquetas de los datos conocidos: $\hat{Y}_l = Y_l$
 - ✓ Etiquetar **el resto** de manera de minimizar: $\hat{Y}^T L \hat{Y}$
- Estrategia 1B:
 - ✓ No forzar las etiquetas de los datos conocidos
 - ✓ Buscar el etiquetado que minimice $\mathcal{H}(\hat{Y})$

La estrategia 1B es preferible para lidiar con ruido en los datos de entrada

Regularización en grafos

Se trata de encontrar el etiquetado \hat{Y} que minimice $\mathcal{H}(\hat{Y})$

$$\mathcal{H}(\hat{Y}) = \|\hat{Y}_l - Y_l\|^2 + \mu \hat{Y}^T L \hat{Y} + \mu \epsilon \|\hat{Y}\|^2$$

$$\|S\hat{Y} - SY\|^2$$

$S = I_{[l]}$ matriz diagonal con 1's en los primeros l elementos y 0 el resto

$$\frac{1}{2} \frac{\partial \mathcal{H}(\hat{Y})}{\partial \hat{Y}} = S(\hat{Y} - Y) + \mu L \hat{Y} + \mu \epsilon \hat{Y} = (S + \mu L + \mu \epsilon I) \hat{Y} - SY = 0$$

$$\frac{1}{2} \frac{\partial^2 \mathcal{H}(\hat{Y})}{\partial \hat{Y} \partial \hat{Y}^T} = S + \mu L + \mu \epsilon I$$

Matriz definida positiva si $\epsilon > 0$ por lo que el costo se minimiza cuando la derivada se anula

$$\Rightarrow \hat{Y} = (S + \mu L + \mu \epsilon I)^{-1} SY$$

- Las etiquetas *optimas* pueden obtenerse invirtiendo una matriz
- Esta matriz depende únicamente del Laplaciano de la red (no de las etiquetas)
- La manera en que las etiquetas se propagan depende exclusivamente de la estructura del grafo

Vínculo entre regularización y propagación de etiquetas

$$\mathcal{H}(\hat{Y}) = \|S\hat{Y} - SY\|^2 + \mu\hat{Y}^T L\hat{Y} + \mu\epsilon\|\hat{Y}\|^2$$

$$\Rightarrow \hat{Y}^* = (S + \mu L + \mu\epsilon I)^{-1}SY$$

Metodo iterativo de Jacobi para invertir una matriz:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

⋮

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n$$

$$x_1 = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n)$$

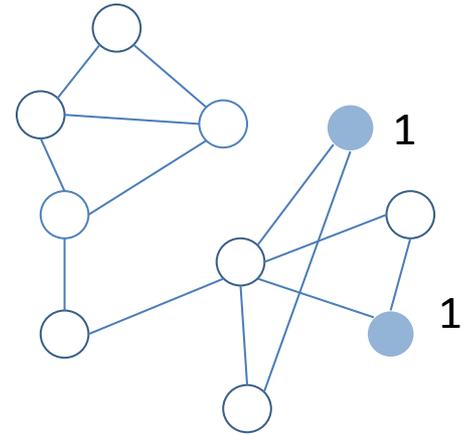
$$x_2 = \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3 - \dots - a_{2n}x_n)$$

⋮

$$x_n = \frac{1}{a_{nn}}(b_n - a_{n1}x_1 - a_{n2}x_2 - \dots - a_{n,n-1}x_{n-1})$$

Metodo iterativo

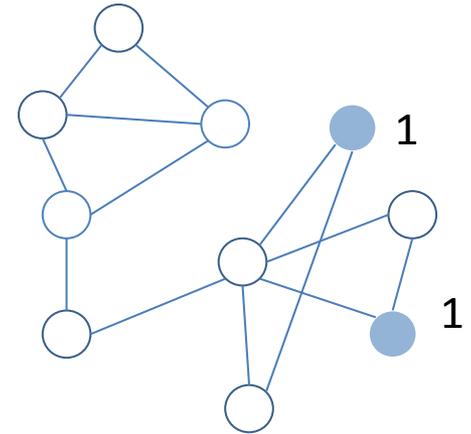
$$x_i^{(t+1)} = \frac{1}{A_{ii}} \left(b - \sum_{j \neq i} A_{ij} x_j^{(t)} \right)$$



Vínculo entre regularización y propagación de etiquetas

$$\mathcal{H}(\hat{Y}) = \|S\hat{Y} - SY\|^2 + \mu\hat{Y}^T L\hat{Y} + \mu\epsilon\|\hat{Y}\|^2$$

$$\Rightarrow \hat{Y}^* = (S + \mu L + \mu\epsilon I)^{-1}SY$$



Metodo iterativo de Jacobi para invertir una matriz:

$$Ax = b$$

$$x_i^{(t+1)} = \frac{1}{A_{ii}} \left(b - \sum_{j \neq i} A_{ij} x_j^{(t)} \right)$$

En nuestro caso $x \equiv \hat{Y}$, $b \equiv SY$, $A = S + \mu L + \mu\epsilon I$ y la iteración de Jacobi resulta en

para nodo etiquetado

$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j \mathbf{W}_{ij} \hat{y}_j^{(t)} + \frac{1}{\mu} y_i}{\sum_j \mathbf{W}_{ij} + \frac{1}{\mu} + \epsilon}$$

para nodo sin etiquetar

$$\hat{y}_i^{(t+1)} \leftarrow \frac{\sum_j \mathbf{W}_{ij} \hat{y}_j^{(t)}}{\sum_j \mathbf{W}_{ij} + \epsilon}$$

Algoritmo de propagación de etiquetas

Regularización en grafos

Hay otras elecciones para la función a minimizar:

$$\mathcal{H}(\hat{Y}) = \|S\hat{Y} - SY\|^2 + \mu \hat{Y}^T L \hat{Y} + \mu \epsilon \|\hat{Y}\|^2$$

$$\mathcal{H}'(\hat{Y}) = \underbrace{\|\hat{Y} - SY\|^2}_{\|\hat{Y}_l - Y_l\|^2 + \|\hat{Y}_u\|^2} + \frac{\mu}{2} \sum_{i,j} w_{ij} \left(\frac{\hat{y}_i}{\sqrt{D_{ii}}} - \frac{\hat{y}_j}{\sqrt{D_{jj}}} \right)$$

Zhou 2004

Dos diferencias entre $\mathcal{H}(\hat{Y})$ y $\mathcal{H}'(\hat{Y})$:

- I. El primer término busca ajustar bien las etiquetas conocidas pero **además** favorece etiquetas 0 para nodos no-etiquetados inicialmente (funciona como regularización)
- II. Las etiquetas están normalizadas por la raíz cuadrada del grado cuando se computa similaridad entre vecinos.

Se puede ver que

$$\mathcal{H}'(\hat{Y}) = \|\hat{Y} - SY\|^2 + \mu (D^{-1/2} \hat{Y})^T L (D^{-1/2} \hat{Y})$$

Regularización en grafos

Hay otras elecciones para la función a minimizar:

$$\mathcal{H}(\hat{Y}) = \|S\hat{Y} - SY\|^2 + \mu \hat{Y}^T L \hat{Y} + \mu \epsilon \|\hat{Y}\|^2$$

$$\mathcal{H}'(\hat{Y}) = \|\hat{Y} - SY\|^2 + \frac{\mu}{2} \sum_{i,j} w_{ij} \left(\frac{\hat{y}_i}{\sqrt{D_{ii}}} - \frac{\hat{y}_j}{\sqrt{D_{jj}}} \right)^2$$

Zhou 2004

$$= \|\hat{Y}_l - Y_l\|^2 + \|\hat{Y}_u\|^2 + \mu \hat{Y}^T (I - \mathcal{L}) \hat{Y} \quad \mathcal{L} \equiv D^{-1/2} W D^{-1/2}$$

$$= \|\hat{Y}_l - Y_l\|^2 + \|\hat{Y}_u\|^2 + \mu \hat{Y}^T D^{-\frac{1}{2}} (D - W) D^{-1/2} \hat{Y}$$

$$= \|\hat{Y}_l - Y_l\|^2 + \|\hat{Y}_u\|^2 + \mu (\hat{Y}^T D^{-\frac{1}{2}}) (D - W) (D^{-\frac{1}{2}} \hat{Y})$$

$$\mathcal{H}'(\hat{Y}) = \|\hat{Y} - SY\|^2 + \mu (D^{-1/2} \hat{Y})^T L (D^{-1/2} \hat{Y})$$

Regularización y propagación de etiquetas

$$\mathcal{H}'(\hat{Y}) = \|\hat{Y} - SY\|^2 + \mu(D^{-1/2}\hat{Y})^T L(D^{-1/2}\hat{Y})$$

$$\frac{1}{2} \frac{\partial \mathcal{H}'(\hat{Y})}{\partial \hat{Y}} = \hat{Y} - SY + \mu(\hat{Y} - \mathcal{L}\hat{Y})$$

$$\Rightarrow \hat{Y}^* = \left((1 + \mu)I - \mu \mathcal{L} \right)^{-1} SY$$

La iteración de Jacobi para resolver $\frac{\partial \mathcal{H}'(\hat{Y})}{\partial \hat{Y}} = 0$

$$\hat{Y}^{(t+1)} \leftarrow \alpha \mathcal{L} \hat{Y}^{(t)} + (1 - \alpha) \hat{Y}^{(0)} \quad \text{con } \mu = \frac{\alpha}{1 - \alpha}$$

O sea..

- ✓ Laplacianos: operador sobre el grafo. Asociados a flujos, random walks, etc...
- ✓ Propiedades espectrales de L
- ✓ Aprendizaje supervisado...busqueda de etiquetados (i.e. campos) **piolas**: conocimiento inicial (parcial), geometria del grafo (smoothness assumption).
- ✓ Label propagation
- ✓ Minimizacion de funcion de costo (cuadratico)
- ✓ Label prop \leftrightarrow minimizacion

Refs

- Bengio, Delalleau, Le Roux: *Label propagation and quadratic criterion.*