

Modelado de Redes Oportunistas

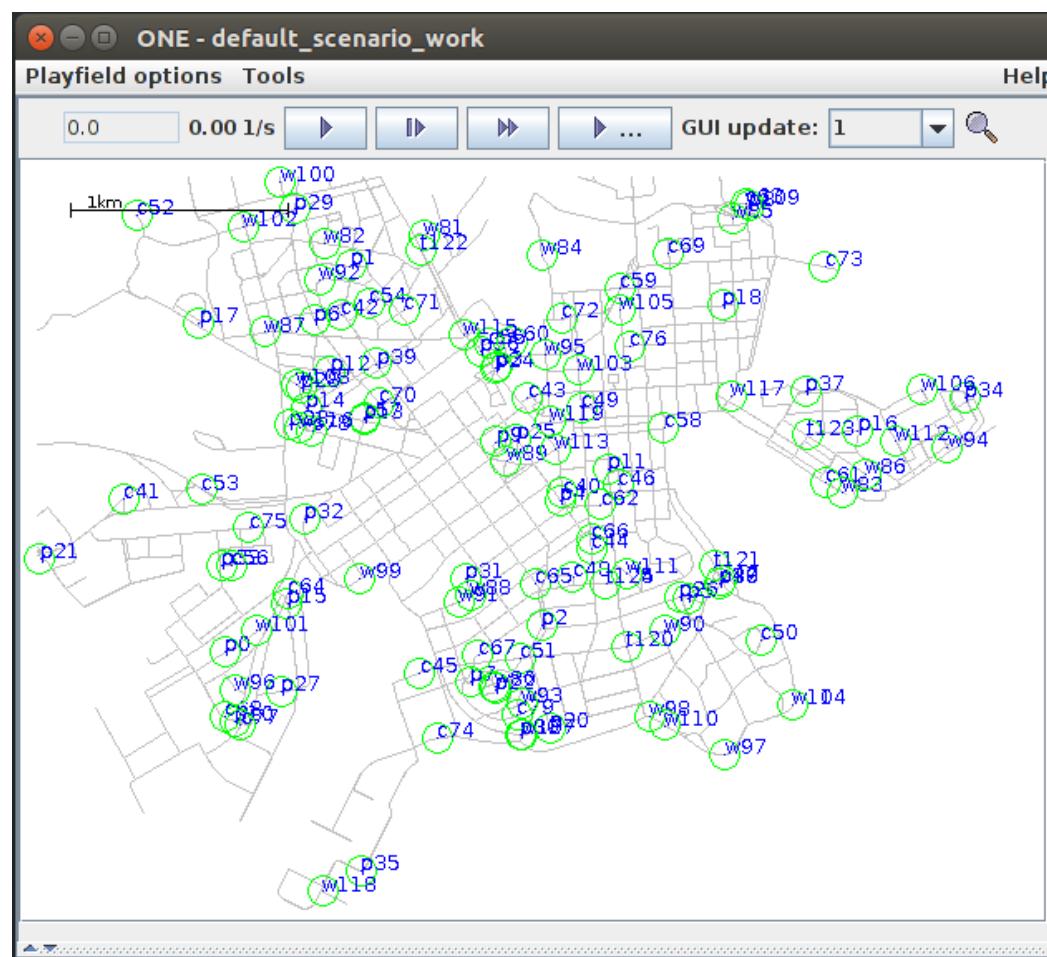
jvisca@fing.edu.uy

Agosto 2019

MINA / Fing-UDELAR

Red oportunista

- Nodos móviles, que pueden intercambiar información cuando se encuentran



Modelos

Analogía con algo que podemos manipular

- Predice, por lo tanto es “útil” o “inútil”
- Deja de ser útil cuando fallan las predicciones: experimento
- No explica, por lo tanto no es “correcto” o “incorrecto”
- Varios modelos distintos pueden ser útiles simultáneamente
- Un modelo puede ser útil “dependiendo de”

Construcción de modelos

"Deductive method of theoretical or axiomatic modeling"

- 1) Planteamos hipótesis
- 2) Deducimos reglas

"Empirical method of experimental modeling"

- 1) Observamos datos
- 2) Inferimos reglas

Construcción de modelos

"Deductive method of theoretical or axiomatic modeling"

- Pro: no menos válido que las hipótesis, siempre
- Con: herramientas limitadas (análisis), hipótesis estrictas

"Empirical method of experimental modeling"

- Pro: ajustado a realidad observada
- Con: ajustado a realidad observada

¿Para qué quiero un modelo?

Quiero predecir:

- ¿Cuál va a ser la tasa de entrega?
- ¿Qué probabilidad hay de que este mensaje llegue?
 - Cuánto va a tardar?
- ¿A cuántos nodos va a llegar este anuncio?
- ¿Qué pasa si
 - mando tanto tráfico?
 - Configuro así el algoritmo?

¿Cómo va a ser mi modelo?

Cosas tiene que representar mi modelo:

- Movilidad de los nodos
- Decisiones que toman los nodos – Algoritmo
 - Enrutamiento / reenvío / recursos
- Generación de tráfico

Deductivo

Buscamos modelos existentes

Primer aproximación: análisis epidémico

- Mensajes son enfermedades
- Copiar es contagiar
- Borrar un mensaje es curarse

Network & traffic model

- N mobile nodes
- Each node carries M messages
- Uniformly distributed
 - Random Direction
 - Constant message arrival rate
- a_x : probability of finding message x in a node
- S : set of active messages

Expected availability

$$\rightarrow Q = \mathbb{E}(a_x) = \frac{M}{|S|}$$

Buffer sizes: static property
↓
 M
↑
Messages alive: dynamics

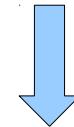
Encounter model

- Nodes $A \rightarrow B$ encounter
- Consider buffers full
- Will copy a message
-infect- if not found on destination
- Will drop *-heal-* if selected for eviction
(random selection)



$$p_{copy}(x, A, B) = p_{x \in A} p_{x \notin B} = a_x (1 - a_x)$$

$$\begin{aligned} p_{evict}(x, A, B) &= p_{x \in B} \frac{1}{M} \#C(A, B) \\ &\approx a_x (1 - Q) \end{aligned}$$



Stochastic logistic epidemic process

“growth of a population in a limited environment”

Resumen:

- Modelo de red: estocástico, uniforme, iid
- Modelo de algoritmo: estocástico, uniforme, iid
- Modelo de tráfico: estocástico, uniforme, iid

Deterministic models

- Take continuous limit for the population
- Copy and eviction form a system of differential equations
- Equating copy and eviction rates, find stable populations:

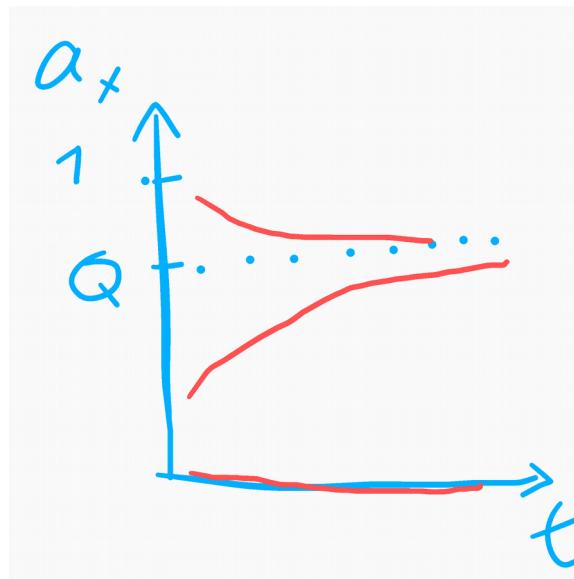
$$\frac{da_x}{dt} = \lambda_N p_{copy} - \lambda_N p_{evict} = \lambda_N a_x(Q - a_x)$$

$$\frac{da_x}{dt} = 0 \Rightarrow a_x = \begin{cases} 0, & \text{repulsive} \\ 1 - \frac{1}{R_0} = Q, & \text{attractive} \end{cases}$$

Deterministic models

Predicción:

- The fraction of nodes with a copy of a message is supposed to settle to an endemic proportion of the population Q .



- But why do messages disappear from the network as new messages arrive?

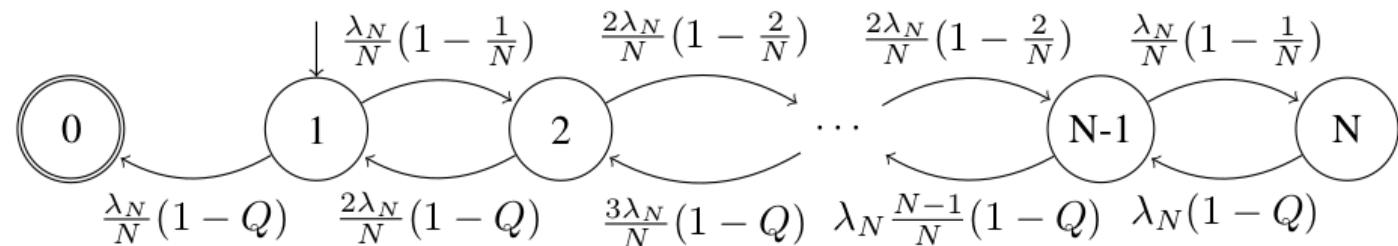
Deterministic models

- Limitación:
 - Determinista a partir de esperanzas.

Stochastic model

Markov chain where states are the number of nodes a message is found on

- Entry state is 1, an absorbing state 0.
- Transitions taken from p_{copy} and p_{evict}
- Messages are dropped when their availability stochastically reach 0

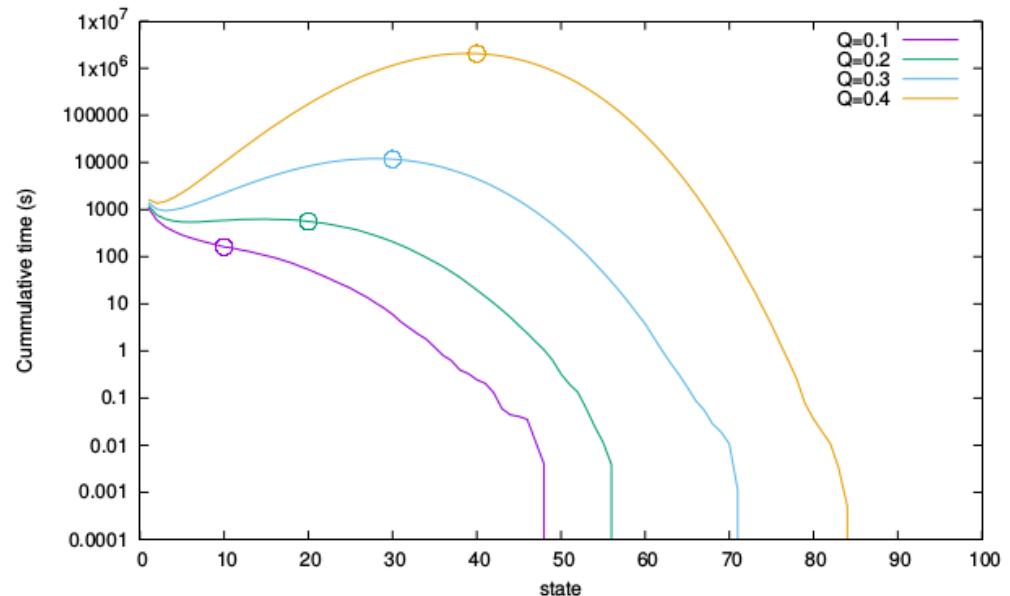


Stochastic model

- Time to absorption from state 1 is the lifetime of a message; using [Norden1981] it is

$$\tau_1 = \frac{N}{\lambda_N} \sum_{k=1}^N \frac{1}{k} \frac{N!}{(N-k)!} \frac{1}{N^k} \frac{1}{(1-Q)^k}$$

- We can compute lifetime of messages as function of Q



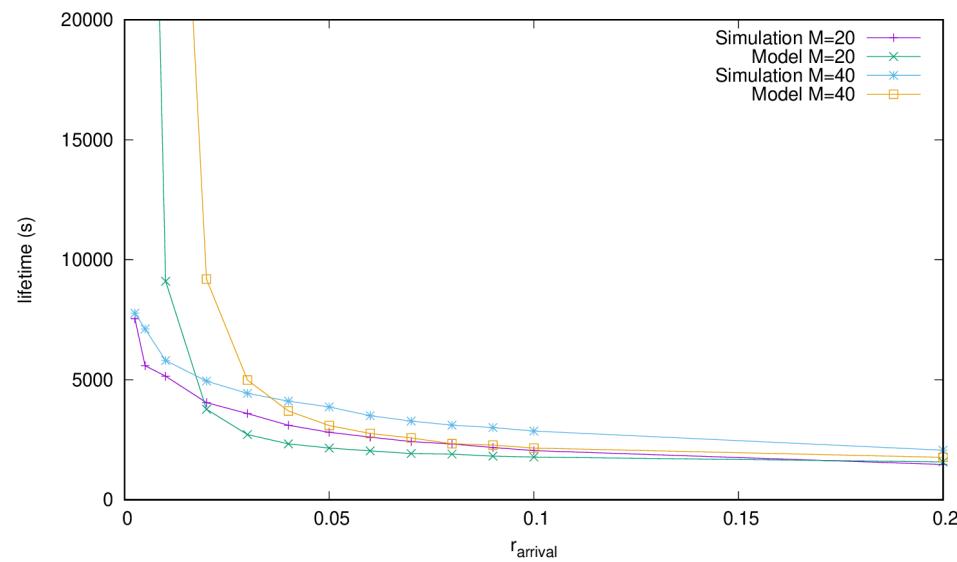
Stochastic model

- Under stable behavior $r_{arrival} = r_{drop}$ we also know:

$$|S| = \frac{B}{Q} = r_{arrival} \times \tau_1(Q)$$

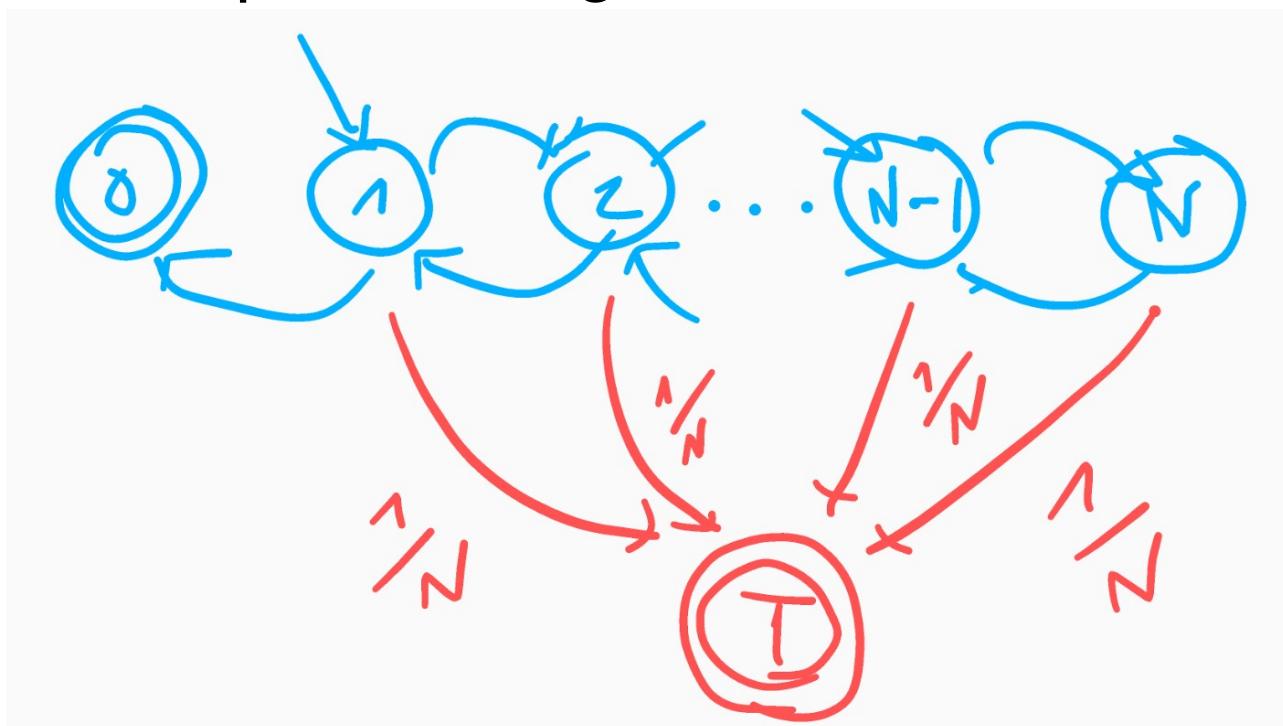
- From these, we can compute $Q=E(a_x)$ as solution of

$$\tau_1(Q) = \frac{B}{r_{arrival}Q}$$



Conclusions

- It is possible to build an analytical model that takes into account
 - Routing algorithm
 - Multiple messages



Pero.

[R5.C1] The authors assume Random Direction mobility model for the nodes in the system. This mobility model is however overly simplified and does not reflect realistic mobility patterns. Yet, the authors claim their findings to help model a realistic network performance. The authors should either justify the choice of mobility model (by providing real-life examples as to where RD is applicable) or show how their model would change with change of the mobility model. In Section 4.5 for instance, nodes are said to traverse straight segments of length 500 m, it is questionable in which real-life scenario this holds.

[R6.C1] The paper discusses a model for epidemic routing under Random Direction mobility. The reference setting is completely unrealistic, even if the authors claim that they develop a model for realistic OPNET. The use of random mobility models has been discouraged for years, since they do not represent the heterogeneity of real human mobility (all nodes are statistically indistinguishable from one another in terms of how they move).

Pero pero.

Nevertheless, we would like to further clarify our approach to the reviewer. Using synthetic random models as scenarios for our proposed model allows us to obtain the model's predictions analytically, parametrized on the same variables of our input mobility model. For example, the computation of the lifetime of messages at Equation 41 receives as inputs: the message arrival rate, buffer sizes and node encounter rate. This allows to study the impact of the different parameters analytically and easily validate against multiple simulated scenarios.

We consider using real data as a next step, where special care must be taken to verify that the scenario properties match the conditions of the model. We feel that applying the model to real data at this point, before validating against synthetic models, would have little predictive value. It would be hard to argue that any good result are not caused by overfitting, correlations or just by chance; conversely, bad results could be caused by some badly understood model requirement being violated by the selected scenario.

Empírico

Datos de la “realidad”

- 12 horas de trazas
- 40 autos
- 120 peatones
- 6 tranvías
- 70m alcance
- Epidemic Routing
- 1Mb buffer, FIFO
- (Mensajes de 10kb,
cada 25-35s, aleatorios, R=0.406)

```
jvisca@fofotimol: ~/workspace/the-one/reports70
jvisca@fofotimol:~/workspace/the-one/reports70$ ls -lh default_
-rw-rw-r-- 1 jvisca jvisca 419M jul 23 16:36 default_scenario_
jvisca@fofotimol:~/workspace/the-one/reports70$ wc -l default_
13079326 default_scenario_work_EventLogReport.txt
jvisca@fofotimol:~/workspace/the-one/reports70$ head default_
0.1 CONN c63 w109 up
0.1 CONN p8 w109 up
0.1 CONN p8 c63 up
0.1 CONN c40 p4 up
0.1 CONN c57 p13 up
0.1 CONN p33 c56 up
0.1 CONN c55 p36 up
0.1 CONN p26 p5 up
0.1 CONN w93 c79 up
0.1 CONN t124 t125 up
jvisca@fofotimol:~/workspace/the-one/reports70$ tail default_
43200.09999965185 S c79 p3 M915
43200.09999965185 DE t124 p0 M1159 R
43200.09999965185 DR p0 M1130
43200.09999965185 S t124 p0 M1366
43200.09999965185 DE c55 c50 M984 R
43200.09999965185 DR c50 M914
43200.09999965185 S c55 c50 M1150
43200.09999965185 DE t121 p21 M1114 R
43200.09999965185 DR p21 M967
43200.09999965185 S t121 p21 M977
jvisca@fofotimol:~/workspace/the-one/reports70$
```

Buscamos una representación

- Modelo para una red:
grafo
- Problema: no tiene en
cuenta secuencia de
encuentros, y
sobreestima la red.

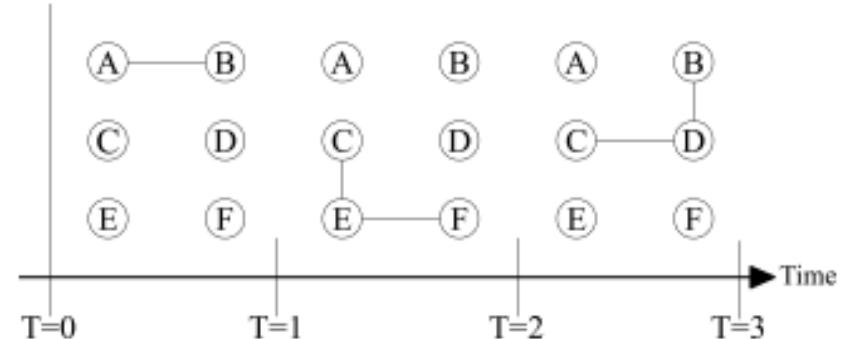


Figure 1: Example Temporal Graph I , $\mathcal{G}_t(0, 3), h = 2$ and $w = 1$.

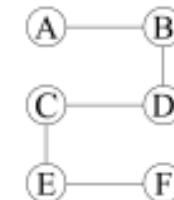


Figure 2: Example static graph based on temporal graph I (Figure 1).

[1] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, “Characterising temporal distance and reachability in mobile and online social networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 1, pp. 118–124, Jan. 2010. [Online]. Available: <http://doi.acm.org/10.1145/1672308.1672329>

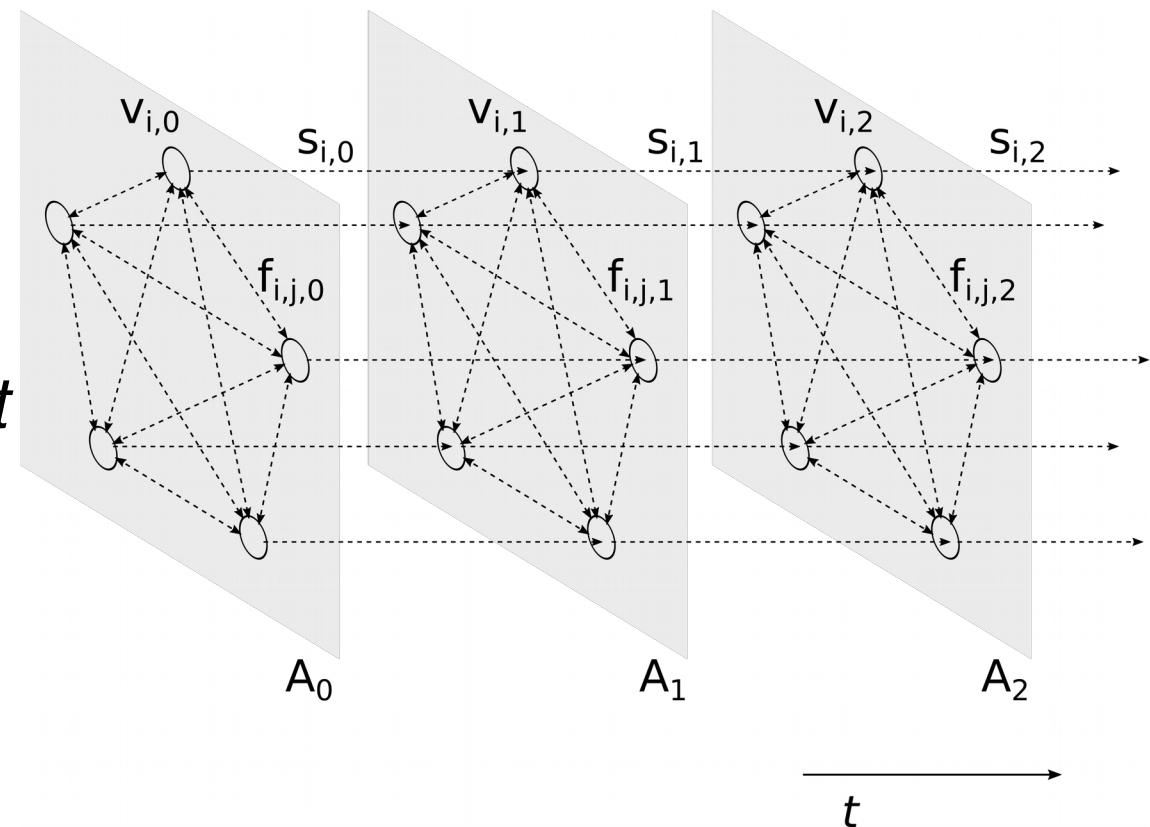
Inventamos representación

Proposed model:

$n_{i,t} \rightarrow$ Agent i in time slot t

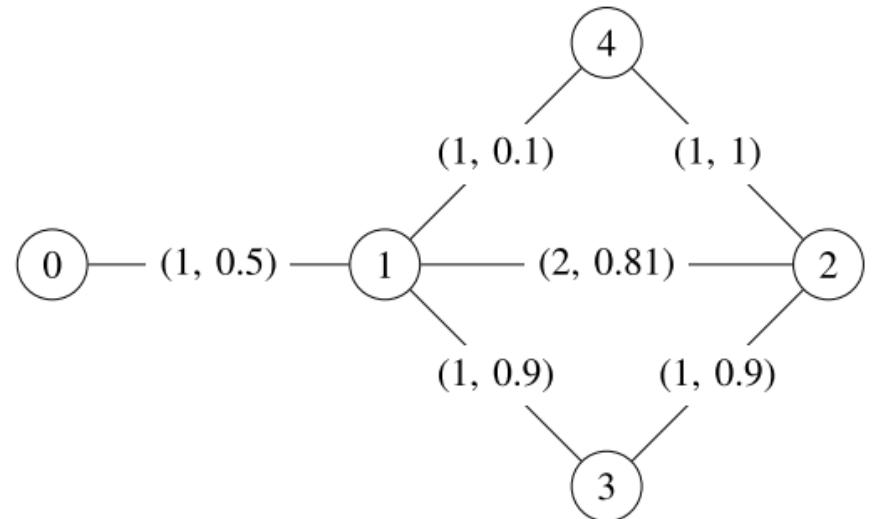
$A_t \rightarrow$ Message transfer
probabilities in time slot t

$s_{i,t} \rightarrow$ Survival probability
for a message in buffer
through time slot t



Interpretación

- Un camino es una trayectoria de un mensaje
- El producto de los enlaces es la probabilidad de recorrerlo



$$P_{path}(E) = \prod_{e_i \in E} p(e)$$

Manipular

Formas de calcular P_{path} :

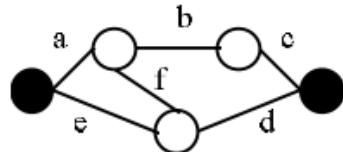
- Dijkstra usando un semiring de probabilidades

name	S	\oplus	\otimes	$\bar{0}$	$\bar{1}$	paths
sp	\mathbb{N}^∞	min	+	∞	0	shortest
bw	\mathbb{N}^∞	max	min	0	∞	widest
rel	[0, 1]	max	\times	0	1	most reliable

- Reemplazando costo $p \leftarrow -\log p$

Interpretar mas

- No es probabilidad de llegar a destino: múltiples caminos, no independientes.
- *Reliability Network* (aka *Flowgraph*):
 - Arista: probabilidad de que enlace está operacional
 - Problema: probabilidad de que haya por lo menos un camino de aristas funcionales
 - NP Completo de los difíciles (*Enumeration Problem*).



The four paths are $\{a,b,c\}$, $\{a,f,d\}$, $\{e,d\}$ and $\{e,f,b,c\}$.

$$\Phi(G) = x_a x_b x_c + x_a x_f x_d + x_e x_d + x_e x_f x_b x_c$$

The six cuts are $\{a,e\}$, $\{a,d,f\}$, $\{b,e,f\}$, $\{b,d\}$, $\{c,d\}$ and $\{c,e,f\}$.

$$\Phi_c(G) = \bar{x}_a \bar{x}_e + \bar{x}_a \bar{x}_d \bar{x}_f + \bar{x}_b \bar{x}_e \bar{x}_f + \bar{x}_b \bar{x}_d + \bar{x}_c \bar{x}_d + \bar{x}_c \bar{x}_e \bar{x}_f$$

Manipular mas

- Soluciones exactas: decenas de nodos.
- Soluciones especiales para topologías particulares:
 - Descomposición Serie/Paralelo:

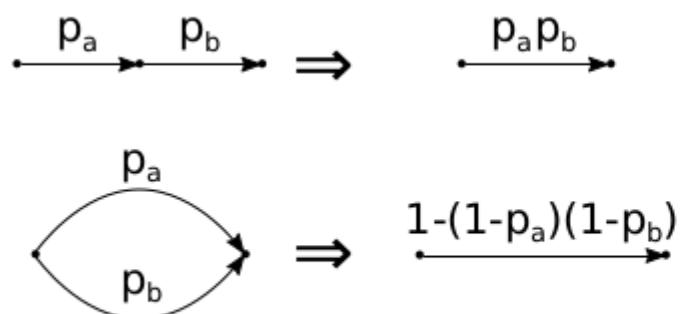


Fig. 3: Serial and Parallel reductions

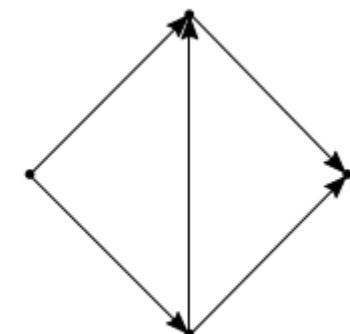
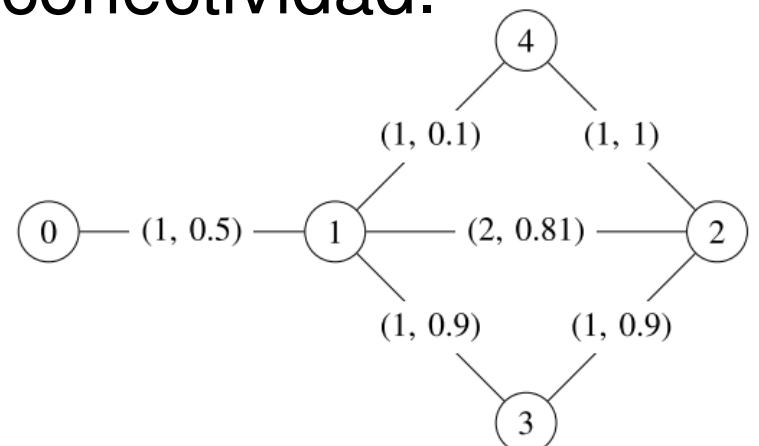


Fig. 4: Irreducible configuration

Manipular mas mas

- Nota: agente representado por múltiples nodos
- Solución viable para redes arbitrarias y grandes: Monte Carlo.
- Versión ingenua:
 - Instancio aristas operativas según probabilidades, verifico conectividad.
 - Repito muchas veces, veo probabilidad



Herramienta útil

Monte Carlo sabio:

1. Genero recorrida DFS desde origen
2. Si visito destino, la recorrida dio *operative*
3. Cuando visito nodo por primera vez, shuffle de vecinos
4. Al seleccionar arista, la sigo o ignoro según Bernoulli de su probabiliad.
5. Si terminé la recorrida, la recorrida dio *failed*.

Herramienta útil

- Bueno: se puede demostrar que la aproximación por Monte Carlo mejora al aumentar los experimentos:

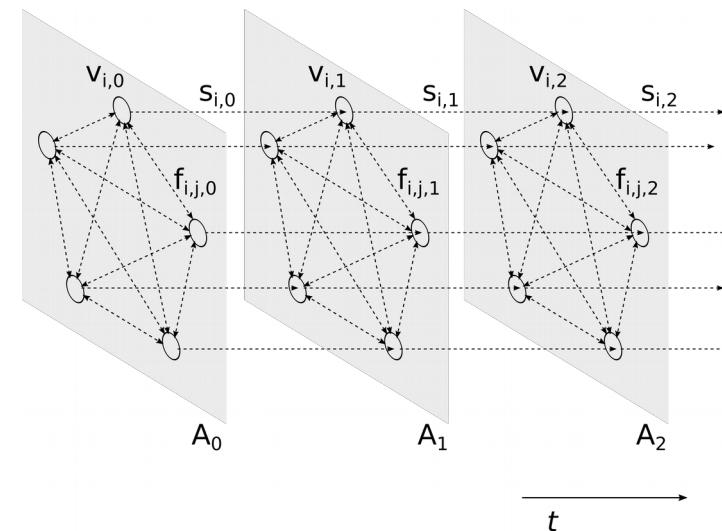
$$\sigma_{\bar{R}} = \sqrt{\frac{\bar{R}(1 - \bar{R})}{r_T}}$$

- Pero el error relativo $\sigma_{\bar{R}}/\bar{R}$ diverge si R se aproxima a 1 o 0 (fenómeno *rare event*)
 - Pero eso no es común en redes oportunistas

Herramienta, aplicación

¿Cómo calcular las probabilidades a partir de las trazas?

- Contando eventos: $p = 1 - \left(1 - \frac{1}{\#B}\right)^D$
- Problema: p es independiente del mensaje
- Realidad: algoritmos tienen estado y miran los mensajes
- Resultado: subestimamos rendimiento



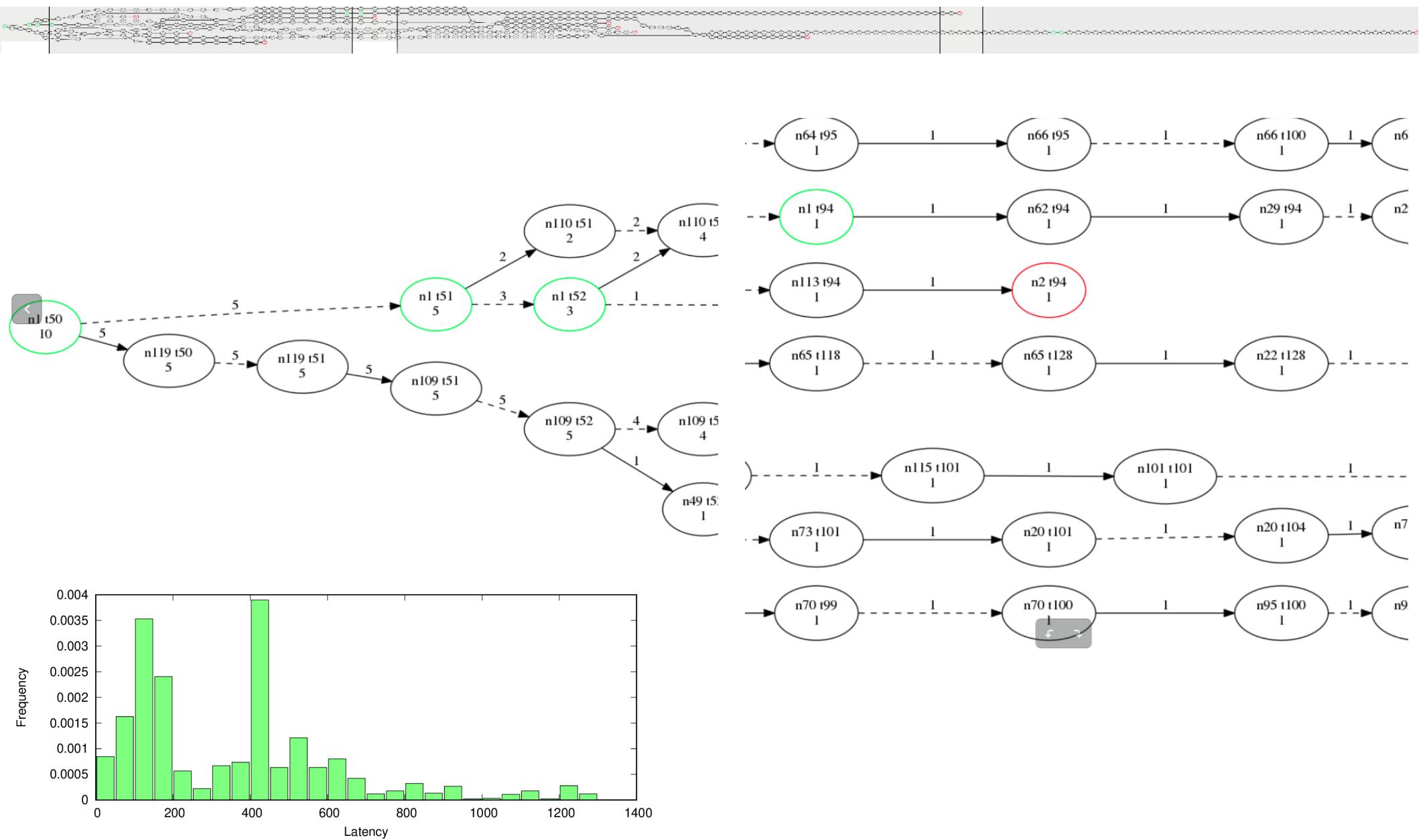
Herramienta, aplicación

Supongamos que podemos representar $p=f(m)$.

¿Cómo construimos f ?

- Solución 1: entrenamos una red neuronal profunda convolutiva (tm).
 - Problema: es por ventana, que puede ser pocos eventos
- Solución 2: la implementamos con conocimiento interno
 - FIFO: estimamos edad de los mensajes dropeados, mantenemos estado contando tiempo en cada agente.

Ejecutamos



¿Cómo usamos?

- Idea: métricas de centralidad / betweenness:
¿Qué tan importante es un nodo para una red?

- Agent centrality,
agente que participa:

$$AC_i = \frac{\sum_{P \in S} In(i, P)}{|S|}$$

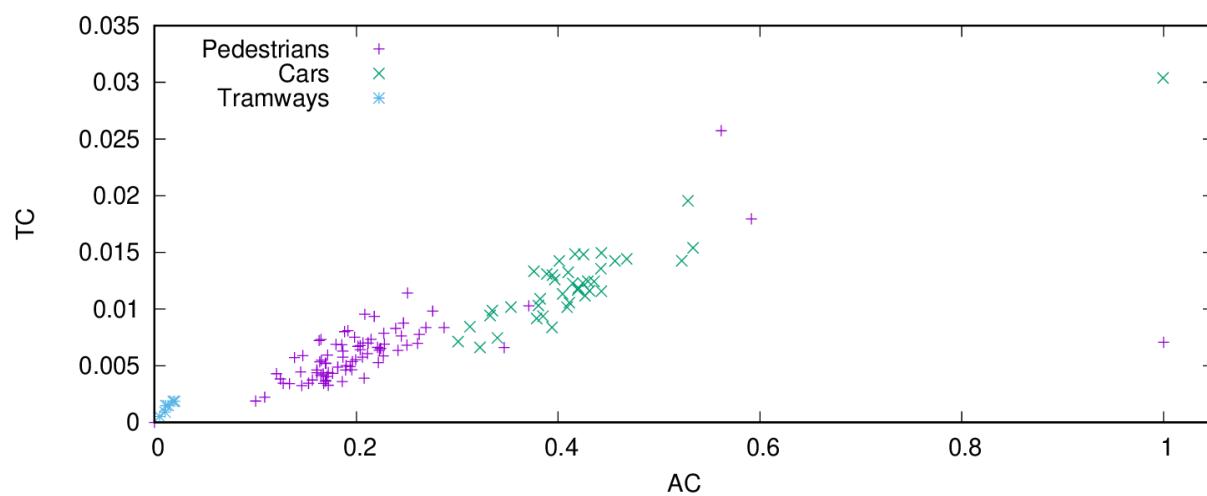
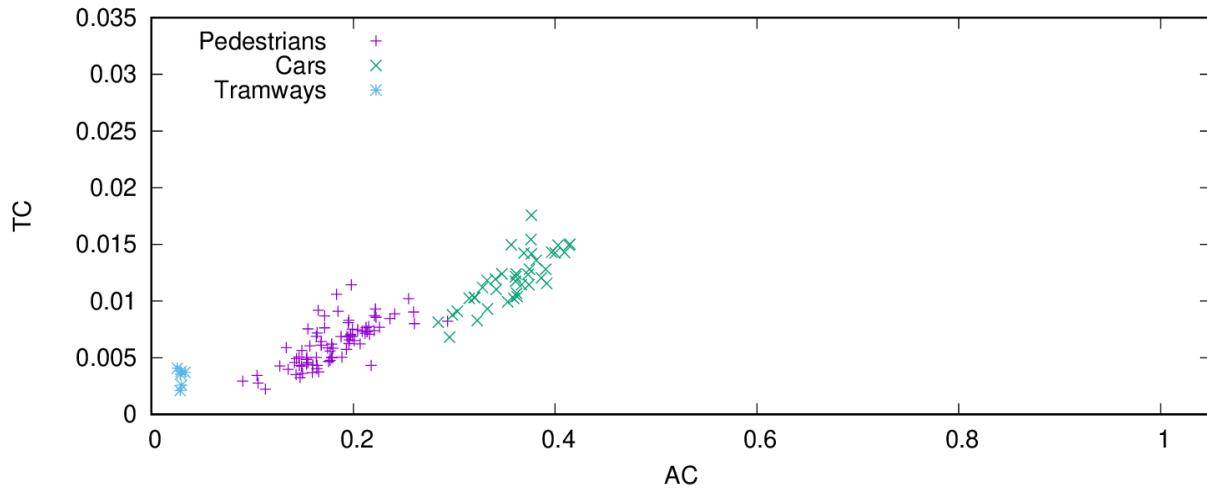
- Time centrality,
agente que invierte tiempo:

$$TC_i = \frac{\sum_{P \in S} |(v_{i,*} \in P)|}{\sum_{P \in S} |P|}$$

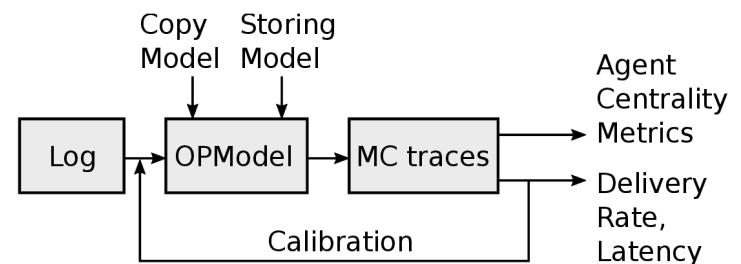
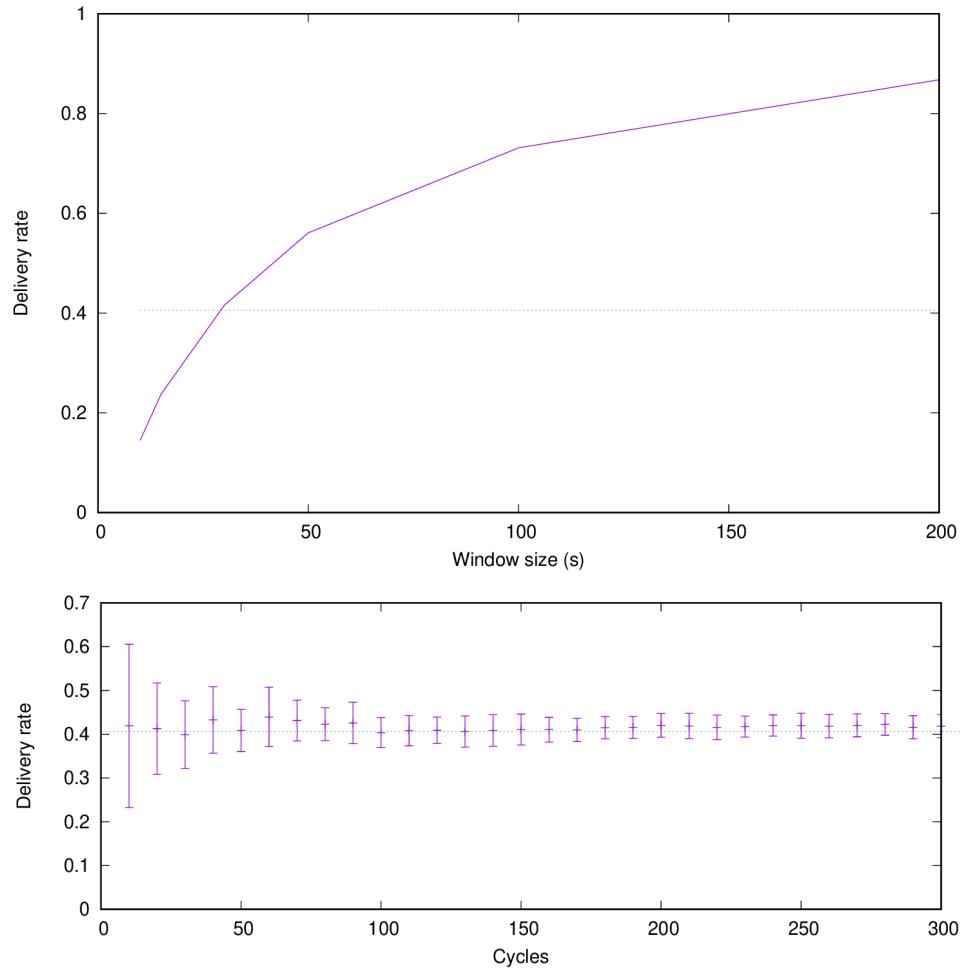
- Transmission centrality,
agente que gasta energía

$$TxC_i = \frac{\sum_{P \in S} |(v_{i,t} \in P, v_{j \neq i, t+1} \in P)|}{\sum_{P \in S} |(v_{j,t} \in P, v_{k \neq j, t+1} \in P)|}$$

Centralidad de agentes



¿Y ahora?



Gracias.