

Protein Interaction Networks: Evolution

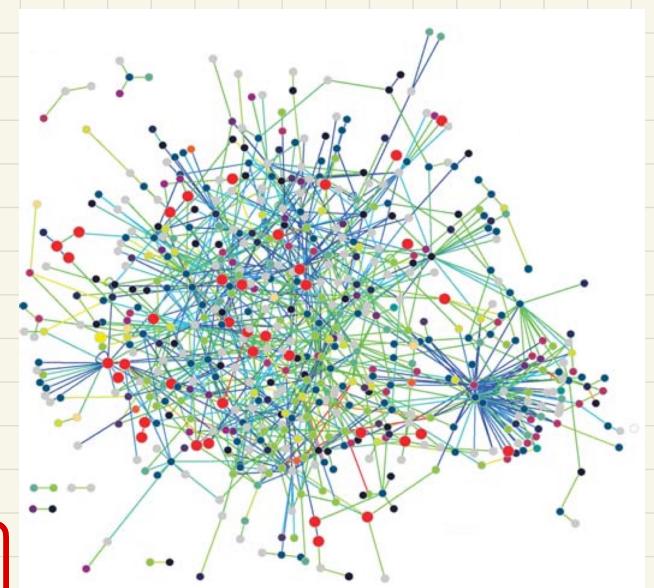
PPINs are complicated!

- multitude of signaling & regulatory pathways
- sense & react to environment
- manage internal processes (growth, division, etc.)

How does evolution generate biological innovation in molecular networks?

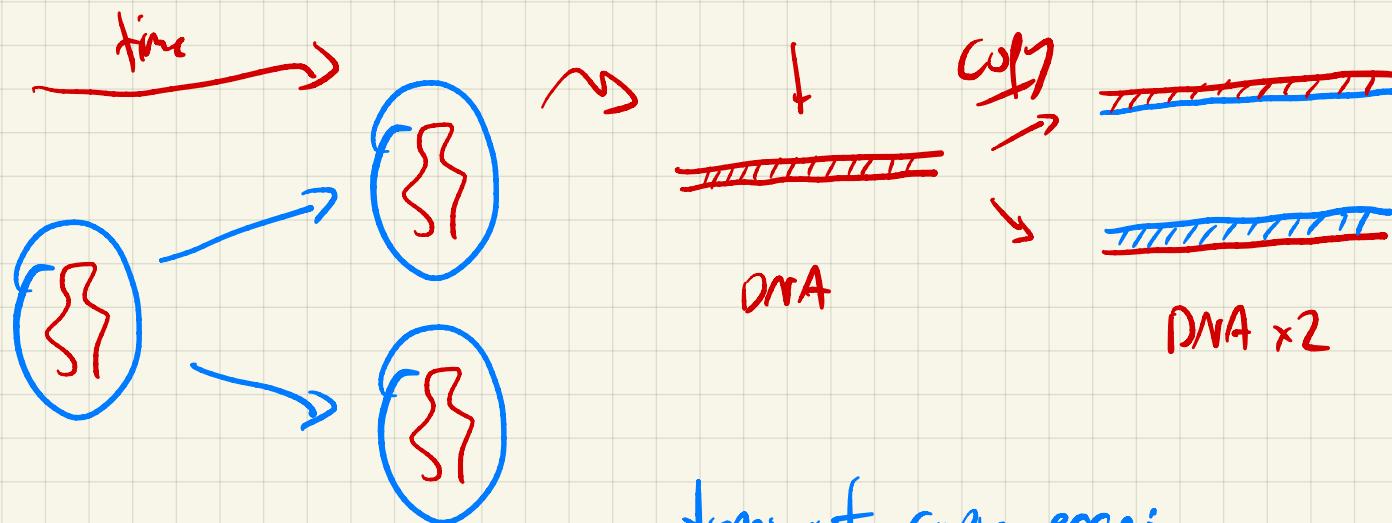
→ molecular lvl: a protein does something new

→ systems lvl: set of proteins (+interactions) do something new



PPIN of *T. pallidum* (bacterium)
[causes syphilis]

How does evolution generate biological innovation in molecular networks?

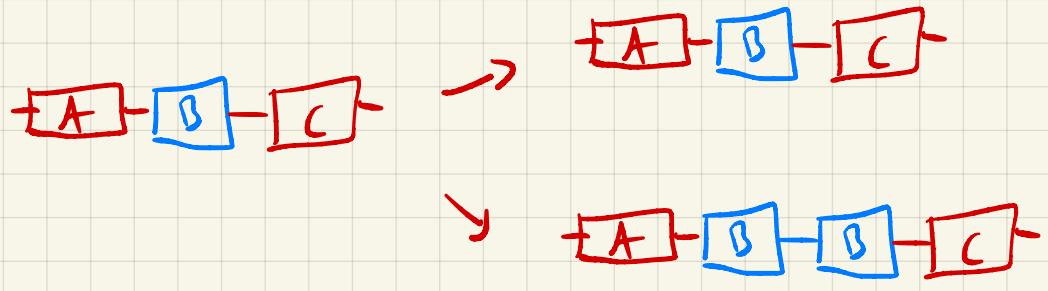


Cell division
→ duplicate entire genome!

* caveat: we consider only protein-coding genes
+ ignore changes in gene regulatory network

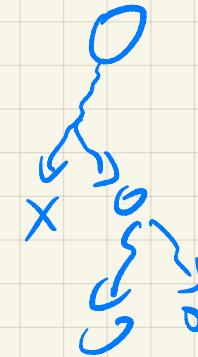
types of copy errors

- 1) SNP - change a letter
- 2) structural variants - duplicate / delete entire sequence



What happens after duplication? \rightarrow selection!

① disadvantageous \rightarrow organism dies



② neutral \rightarrow neither good nor bad

\hookrightarrow random walk
(genetic drift)



③ advantageous \rightarrow selection favors these individuals!



if neutral or advantageous:

(A) Neofunctionalization

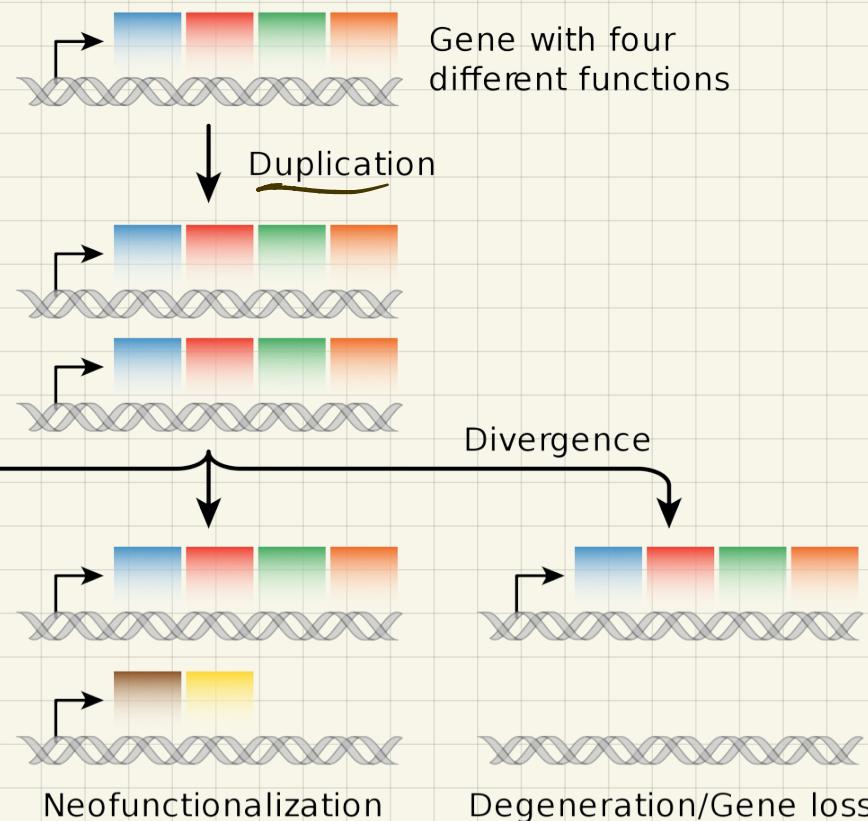
partition functionality

New function

(B) Subfunctionalization

shared functionality

divide



*caveat:
functionality is not
linear / can also
change its strength

Duplication-divergence network models

Nodes: protein coding genes
edges: PPIs

In the network: duplicate a gene = duplicate a node + its links

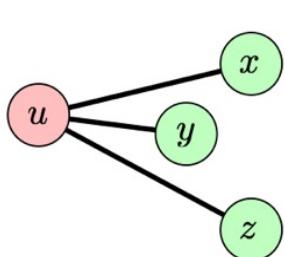
* many variations / operationalizations → many similar models

* key assumption: separation of time scales → gene duplications ~ 10s of generations
edge modifications ~ 100s or
10,000s of gen.

Network model:

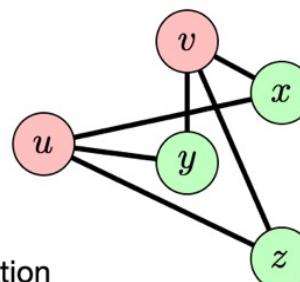
Step 1: choose

$$E = \{(u, x), (u, y), (u, z)\}$$



duplication

Step 2: duplicate



$$E = \{(u, x), (u, y), (u, z), (v, x), (v, y), (v, z)\}$$

optional: complement $\textcircled{v} - \textcircled{u}$

- only nodes with connections can gain new edges (if a neighbor duplicates)
- what happens to G if we only duplicate?

• add in neo/subfunctionalization: divergence \rightarrow edge rewiring

* many ways to operationalize these mechanisms \rightarrow different flavor models

how to rewire: for each duplicated edge (v, x)

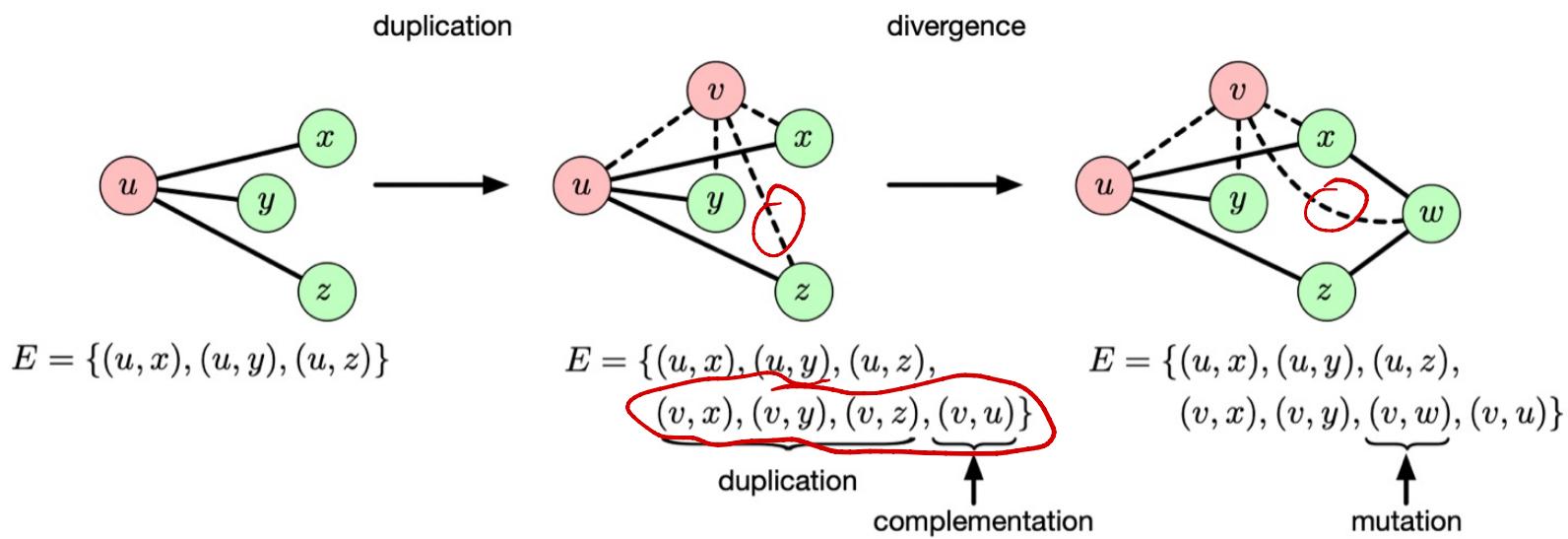
} keep it
 } rewrite as (v, w)
 for uniformly
 random w

with prob q
with prob $1-q$

step 1: choose

step 2: duplicate

step 3: rewire

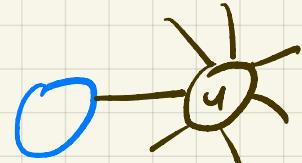


→ growth-only model (no gene loss + no edge rewiring outside duplicates)

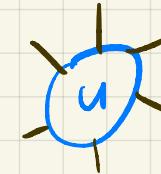
Network measures

- degree distribution : thru ways a node $k_i \rightarrow k_i + 1$

1) a neighbour of i duplicates w/ prob q



2) chose as $n \leftarrow$ for some other rewiring event



neighbor duplication

choose u w/ prob $\frac{1}{n}$

it has degree k_u

prob one of its neighbors duplicate is

$$\frac{k_u}{n} q$$

"uniform attachment"

depends on mean degree of net, c

$c(l-i)$: # of rewired edges

$$\frac{c(l-i)}{n}$$

$$\Pr(k_i \rightarrow k_i + 1) \propto k_i$$

$$\Pr(\underbrace{k_i \rightarrow k_i + 1}) = \frac{k_i q}{n} + \frac{(1-q)c}{n} = \frac{k_i q + (1-q)c}{n}$$

this pattern is called "preferential attachment" (equiv. to Yule process)

\downarrow math!

$$\Pr(k) \propto k^{-\alpha} \quad \text{power-law!} \quad \underline{\alpha = 1 + \frac{1}{q}}$$

a.k.a. Cumulative advantage

→ Near-perfect copying ($q \approx 1$) : $\alpha \approx 1 + \frac{1}{1} = 2$

*caveat:
no deletion!

→ Poor copying ($q \approx 0$) : $\alpha \approx 1 + \frac{1}{0}$

- clustering coefficient \geq configuration model $O(\frac{1}{n})$
↳ duplication tends to "localize" edges

- mean geodesic distance (MGD) $O(\ln n)$ like a random graph

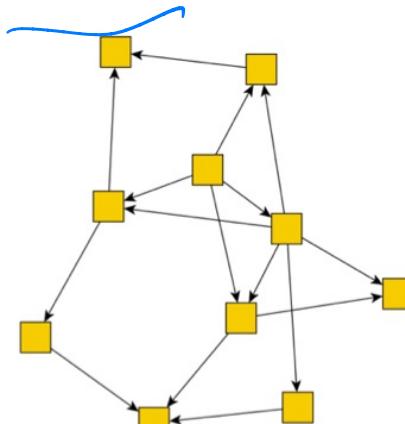
Simulating PPLN evolution (Thursday!)

easy to simulate

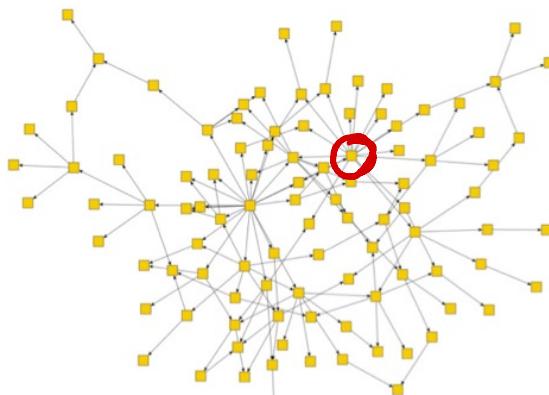
$$\ell = 1/2$$

$$\alpha = 1 + \gamma_{1/2} = 3$$

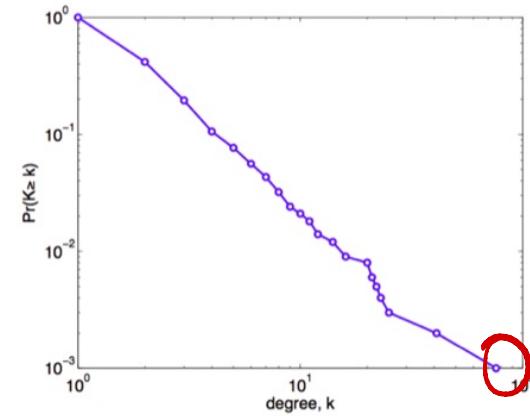
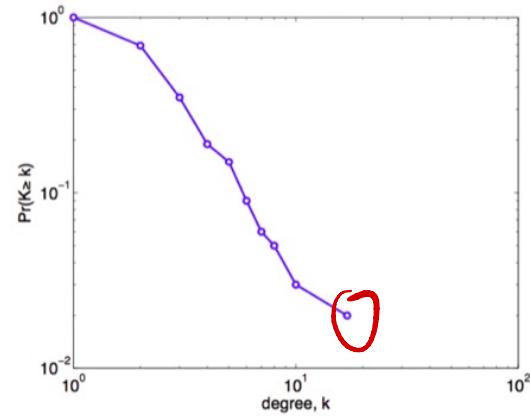
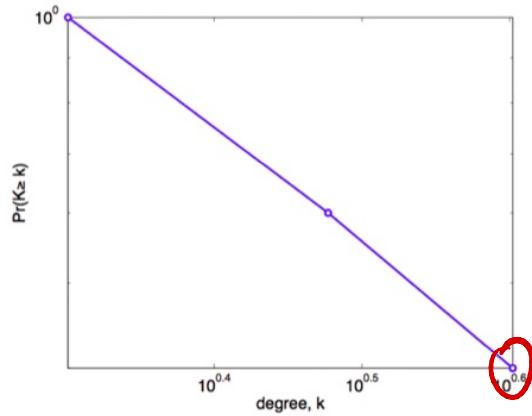
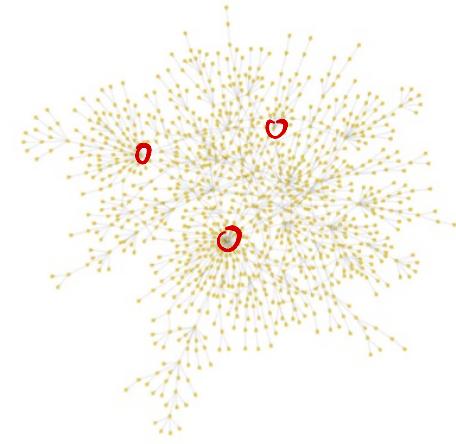
$n = 10$



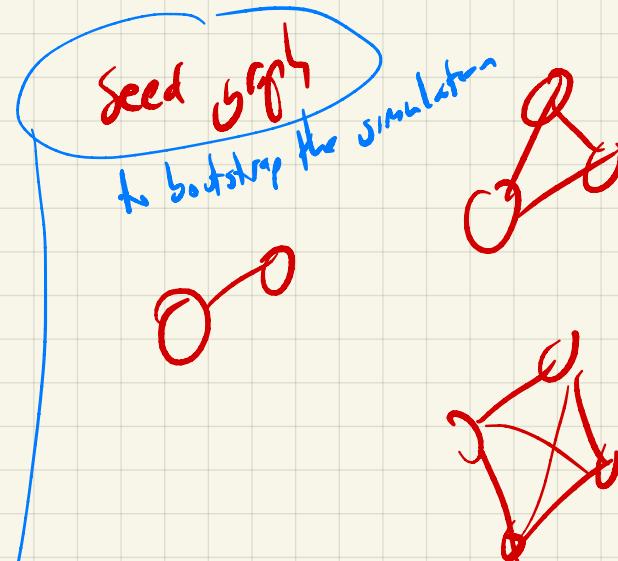
$n = 100$



$n = 1000$



time matters! → What's going on here?



PPIN lab

q : edge modification rate (subfunctionalization)

q_c : complementation rate

flag_neu : use neofunctionalization or not

During each duplication event $\{u\} \rightarrow \{u, v\}$, the DMC model allows for

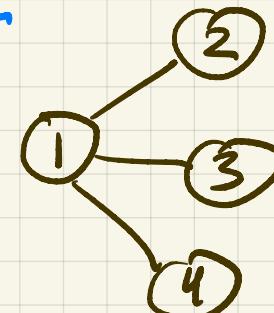
- *subfunctionalization*, in which u and v divide up u 's previous interaction partners,
- *neofunctionalization*, in which v acquires novel interaction partners, and
- *complementation*, in which u and v also become interaction partners.

In pseudocode, here's how the DMC model works:

- (setup) choose q and q_c , and set the number of nodes n at the end
- (setup) define a "seed graph" $G_0 = (V_0, E_0)$ with some initial topology
- at each time step $t < n$ of the simulation
 - (dup) choose a node $u \in V_{t-1}$, uniformly at random
 - (dup) create a new node v
 - for each $(u, x) \in E_{t-1}$
 - (copy) add the edge (v, x) to E_t
 - with probability q (modify a copied edge)
 - (subfun) flip a fair coin, and delete either (u, x) or (v, x)
 - otherwise
 - (neofun) choose y uniformly at random from V_{t-1}
 - (neofun) replace (v, x) with (v, y)
 - (comp) with probability q_c , add the "complement" edge (u, v) to E_t

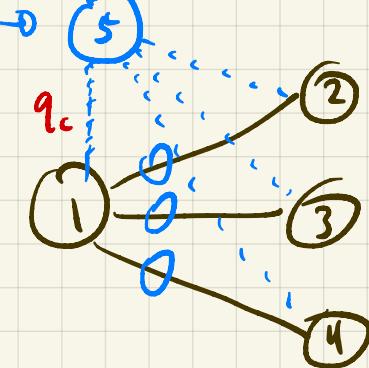
seed graph

G_0 :



G_1

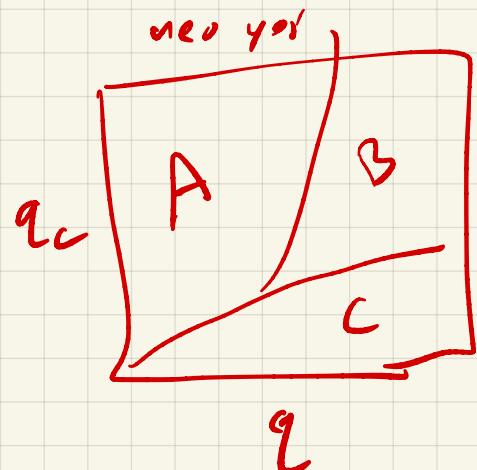
G_1 :



- for each edge $(1, x)$:
 - q : keep either $(1, x)$ or $(5, x)$
 - $1-q$: if flag_neu :
remove $(5, x)$ + $(5, y)$
- q_c : add $(1, 5)$

Part 1,a: Building intuition about q

Explore the effect of q (the edge modification rate) on the kinds of graphs that DMC grows
Try various values of q
Start with small n, then crank it up
Discuss how well this behavior aligns with what we saw in Lecture 10

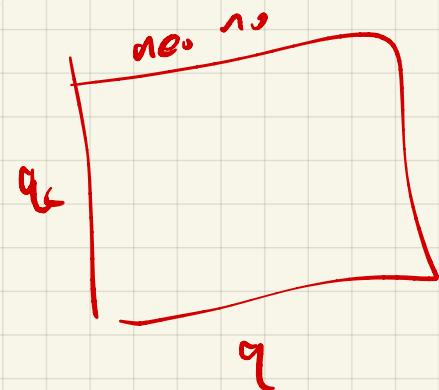


Part 1,b: Building intuition about qc and flag_neo

Explore effects of qc and flag_neo on the kinds of graphs that DMC grows
Discuss with your teammates

Part 1,c: Growing large networks via DMC

Choose 3 values of q that represent qualitatively different types of graphs, grow a n=1000 network
Discuss whether model does or does not perform as expected; if not, develop ideas about why



Part 1,d: Quantifying the shape of growing DMC networks

Augment your simulation to calculate $\langle k \rangle$, $\langle \text{ell} \rangle$, C and make some nice figures
Discuss what these patterns tell you about how the graph's structure is evolving

Part 2,a: What's the difference?

Using the directed duplication-divergence sim, grow some networks of sizes n=10,50,1000
Explore effect of q on shape of network and final $\text{Pr}(k)$
Discuss differences you see with the DMC model from Part 1