# Exploring a Graph Using BFS

15-451
9/1/20

There are at least 3 methods to explore a graph:

1) DFS (1970's)
2) BFS 1980, 2000, 2010's (FRT trees)
3) Random Walks
   (Critical for modern alg design)

# Low Diameter Decomposition

Applications of Low Dia Decomp

1) Spanners (Distance Preserving Sparse Graphs)
2) Hop Set (Added edges to decrease number of edges needed in shortest paths)
3) Low Stretch Spanning Tree (LSST)

(Trees preserve distances on average)

Applications of LSST

Fast Alg for
1) Linear Solvers
2) Max Flow
3) Image Processing

4) Metric embeddings

(Emedding one metric into another)

# Low Dia Decomp

Let $G=(V,E)$ undirected (unweighted)

Def Boundry Edges$(W) \equiv$

$$\partial W = \{(x,y) \mid x\in W,\ y\notin W,\ (x,y)\in E\}$$

Internal Edges $\equiv$

$$I(W)=\{(x,y) \mid x\in W,\ y\in W,\ (x,y)\in E\}$$

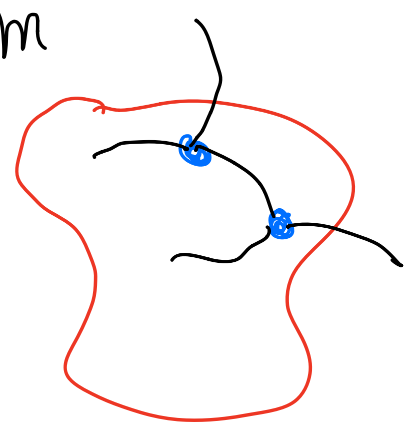
$d(v) \equiv$ degree of $v \in V$

---

Def: $\mathrm{Vol}(W) \equiv \sum_{v\in W} d(v) \quad W \subseteq V.$

Note: $\mathrm{Vol}(W) = 2|I(W)| + |\partial W|$

Thus: $\mathrm{Vol}(V) = 2m$



# Finding a low diameter set



**Def** Isoperimetric Number $(W) \equiv$

$$\Phi(W) = \frac{|\partial W|}{\mathrm{Vol}(W)}$$

---

Prob: Given $G=(V,E)$, $x \in V$, $0<\beta<1$

Find $x \in W \subseteq V$ of nearby points

st $\Phi(W) \leq \beta$.

We will use BFS.

## Ball Growing

**Def** $B(x,r) = \{ y \in V \mid dist(x,y) \le r \}$

(Ball of radius $r$ centered at $x$)

Procedure: GrowBall$(G, x, \beta)$

Set $r = 1$

While $\Phi(B(x,r)) > \beta$ set $r = r+1$

Return $B_r = B(x,r)$, $R = r$

---

# GrowBall Analysis

**Claim:** $R = O(\log n / \beta)$

note: If $r < R$ then $|\partial B_r| \geq \beta(\mathrm{Vol}(B_r))$

Since $\frac{|\partial B_r|}{\mathrm{Vol}(B_r)} > \beta$

---

**Def** $\overline{\partial(B_r)} = \{ y \mid (x,y) \in E, x \in B_r, y \notin B_r \}$

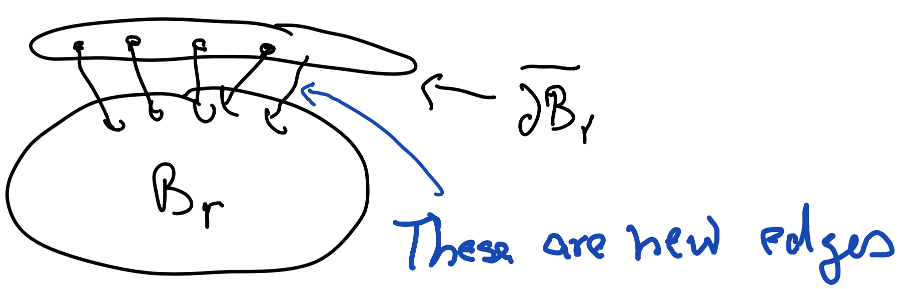
$\overline{\partial(B_r)} = B_{r+1} \setminus B_r$

---

**Claim:** $\mathrm{Vol}(\overline{\partial(B_r)}) \geq \beta \mathrm{Vol}(B_r)$

**thus** $\mathrm{Vol}(B_{r+1}) \geq (1+\beta)\mathrm{Vol}(B_r)$

**Pf:**



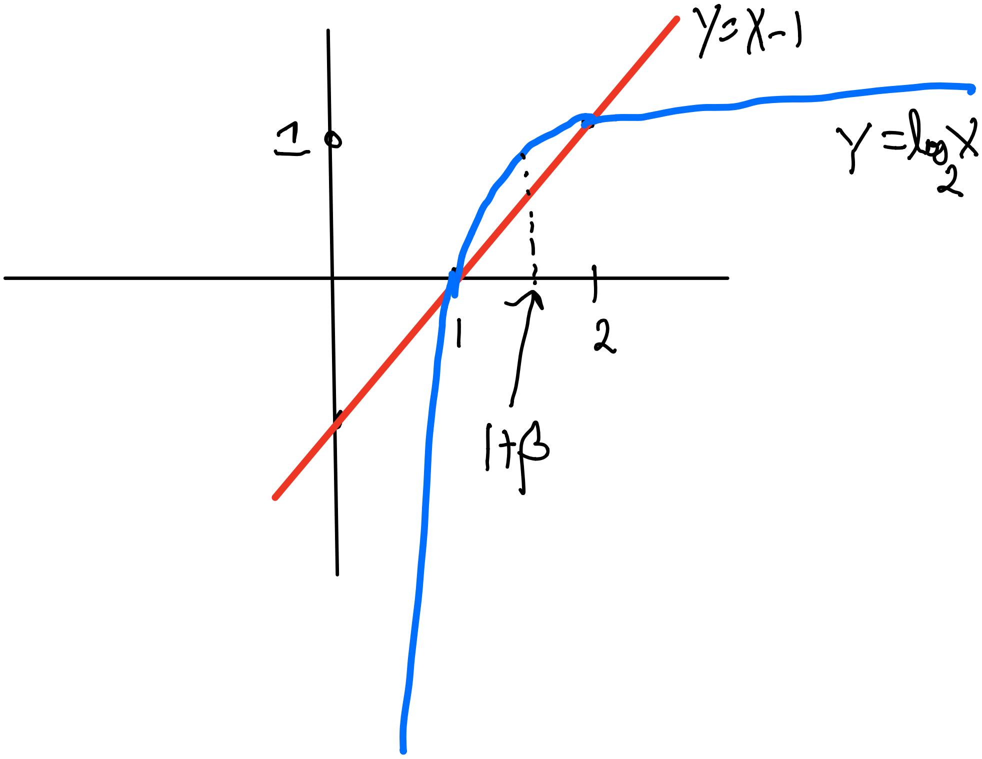
**Thus:** $(1+\beta)^r \leq \mathrm{Vol}(B_r) \leq 2m \quad *$

**Claim:** $r \leq \dfrac{\log_2 m + 1}{\beta}$ for $0 < \beta \leq 1$

**Pf:** Taking logs of $*$

$$r \log_2(1+\beta) \leq \log_2 2m$$

**Subclaim:** $\log_2(1+\beta) \geq \beta \quad 0 \leq \beta \leq 1$



<u>To show</u>: $f(x) = \log_2(x) - x + 1 \geq 0$ for $1 \leq x \leq 2$

Note $\log_2(x) = x - 1$ for $x = 1, 2$

$\log_2 x = \log_2 e \ln x$

$$f'(x) = \frac{\log_2 e}{x} - 1$$

$$\& \; f''(x) = -\frac{\log_2 e}{x^2}$$

<u>Thus</u>: $f$ is a concave fcn.

ie $f(\alpha a + \beta b) \geq \alpha f(a) + \beta f(b)$ $\alpha + \beta = 1$
$\alpha, \beta \geq 0$

---

$$r \cdot \beta \leq \log_2 m + 1 \Rightarrow r \leq \frac{\log_2 m + 1}{\beta}$$

We now use GrowBall to get a partition of $V$.

Procedure: BallDecomp$(G, \beta)$

While $V \neq \emptyset$

1) pick $x \in V$

2) $B_r = \text{GrowBall}(G, x, \beta)$

3) Remove $B_r$ & $\partial B_r$ from $G$

Return Balls.

---

<u>Claim</u>: # interball edges (cut) $\leq 2\beta m$

<u>Pf</u> Suppose BallDecomp returns balls $B_1, \dots, B_k$

1) with volumes $Vol_1, \dots, Vol_k$

2) bdry sizes $\partial_1, \dots, \partial_k$ (At the time removed)

Then $\sum Vol_i = 2m$, $\sum \partial_i = \text{cut}$, and $\beta Vol_i \geq \partial_i$.

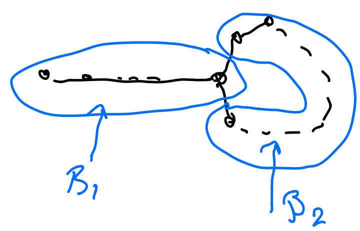
<u>Thus</u>: $2\beta m = \beta \sum Vol_i \geq \sum \partial_i = \text{cut}$

$\therefore 2\beta m \geq \text{cut}$ $\square$

<u>Note</u>: $\mathrm{dist}_G(v,w) \ll \mathrm{dist}_B(v,w)$ for

$v,w \in B \equiv$ Ball

eg



# Ball Growing Using Exponential Delay

Procedure: ExpDelay$(G, \beta)$

1) Each vertex $v \in V$ draws $X_v \sim Exp(\beta)$
2) Each $v \in V$ computes $S_v = X_{max} - X_v$
3) Each $v \in V$ starts a BFS at time $S_v$
   a) If $v$ is not owned at time $S_v$ then $v$ owns $v$.
   b) Each $v$ is owned by first arrival vertex.

---

<u>Def</u> $u \in cluster(v)$ if

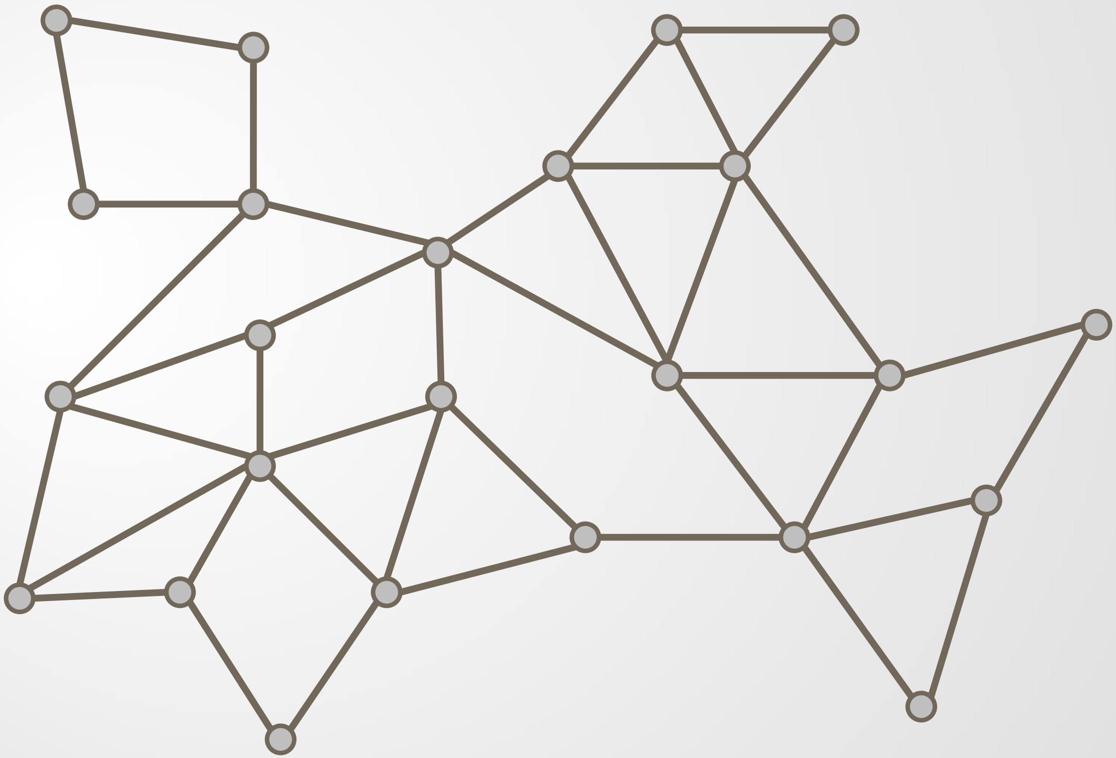
1) $v = \underset{w}{\arg\min} \{ dist(w,u) + S_w \}$

   or equivalently:

2) $v = \underset{w}{\arg\max} \{ X_w - dist(w,u) \}$

# An Example Run [MPX13]

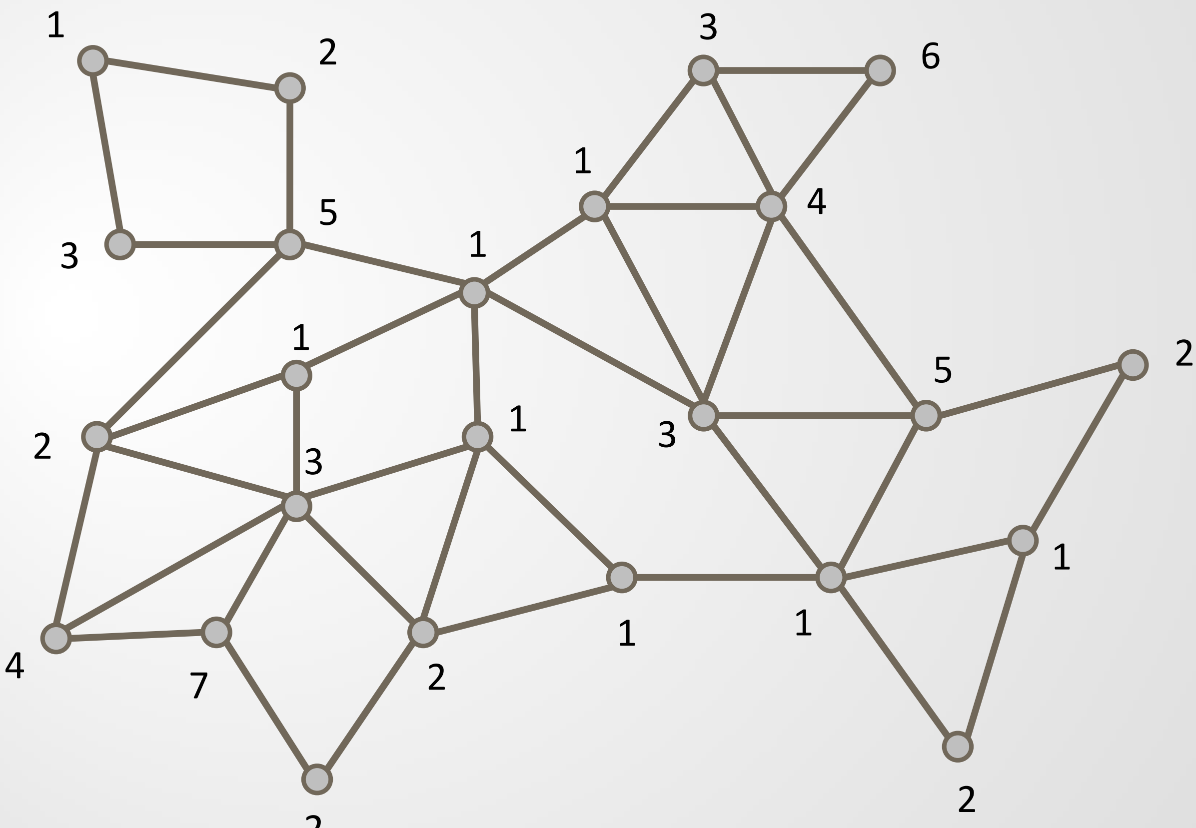
A



Vertex $v$ draws $X_v \sim \mathrm{Exp}(\beta)$

$\beta$

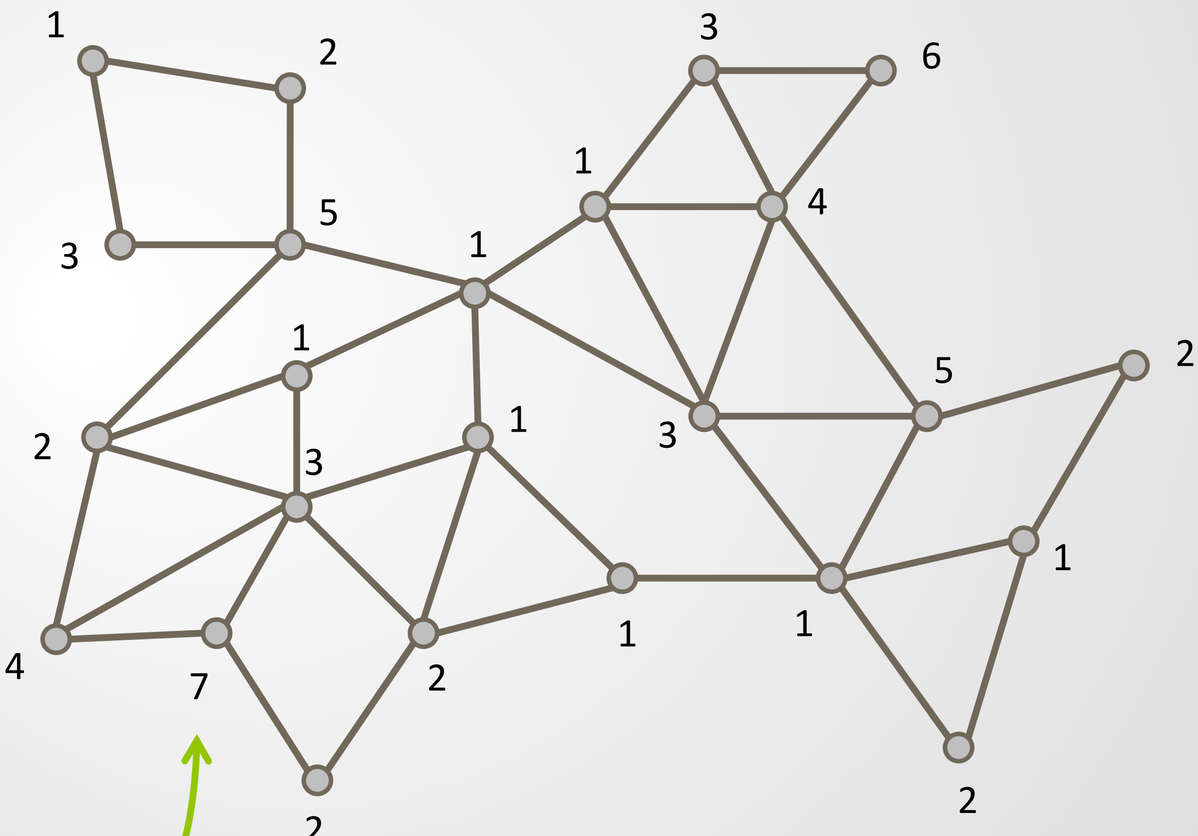
# An Example Run [MPX13]




Vertex $v$ draws $X_v \sim \mathrm{Exp}(\beta)$

C

# An Example Run [MPX13]

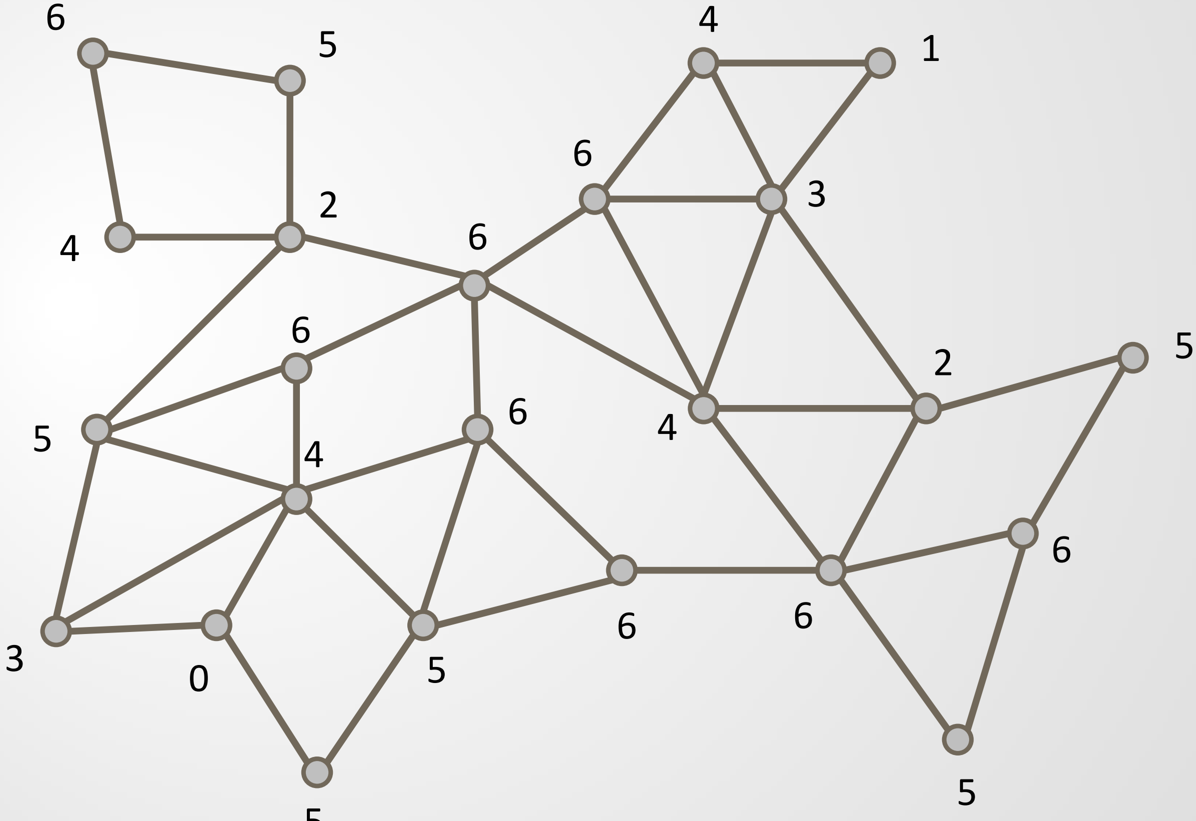



$X_{max} = 7$

Vertex $v$ computes $X_{max} - X_v$

D

# An Example Run [MPX13]

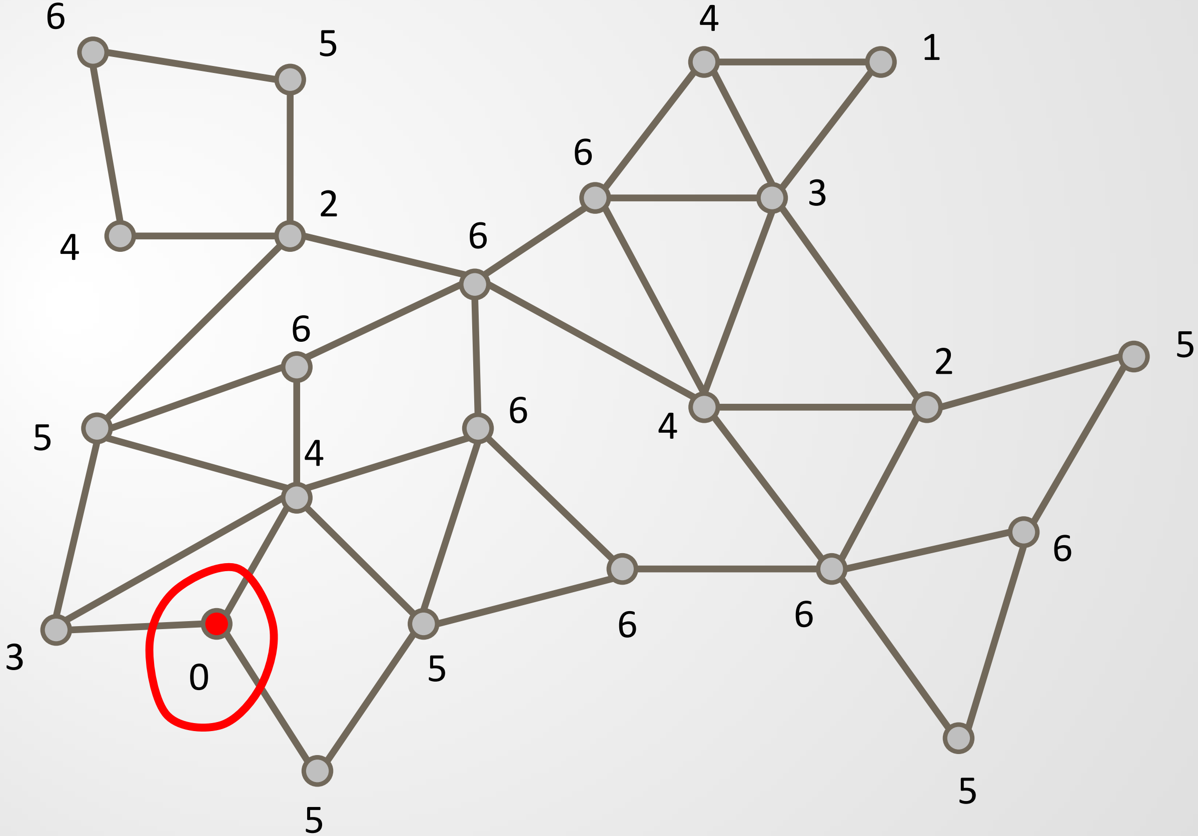



Vertex $v$ computes $X_{max} - X_v$

# An Example Run [MPX13]

$\mathcal{E}$

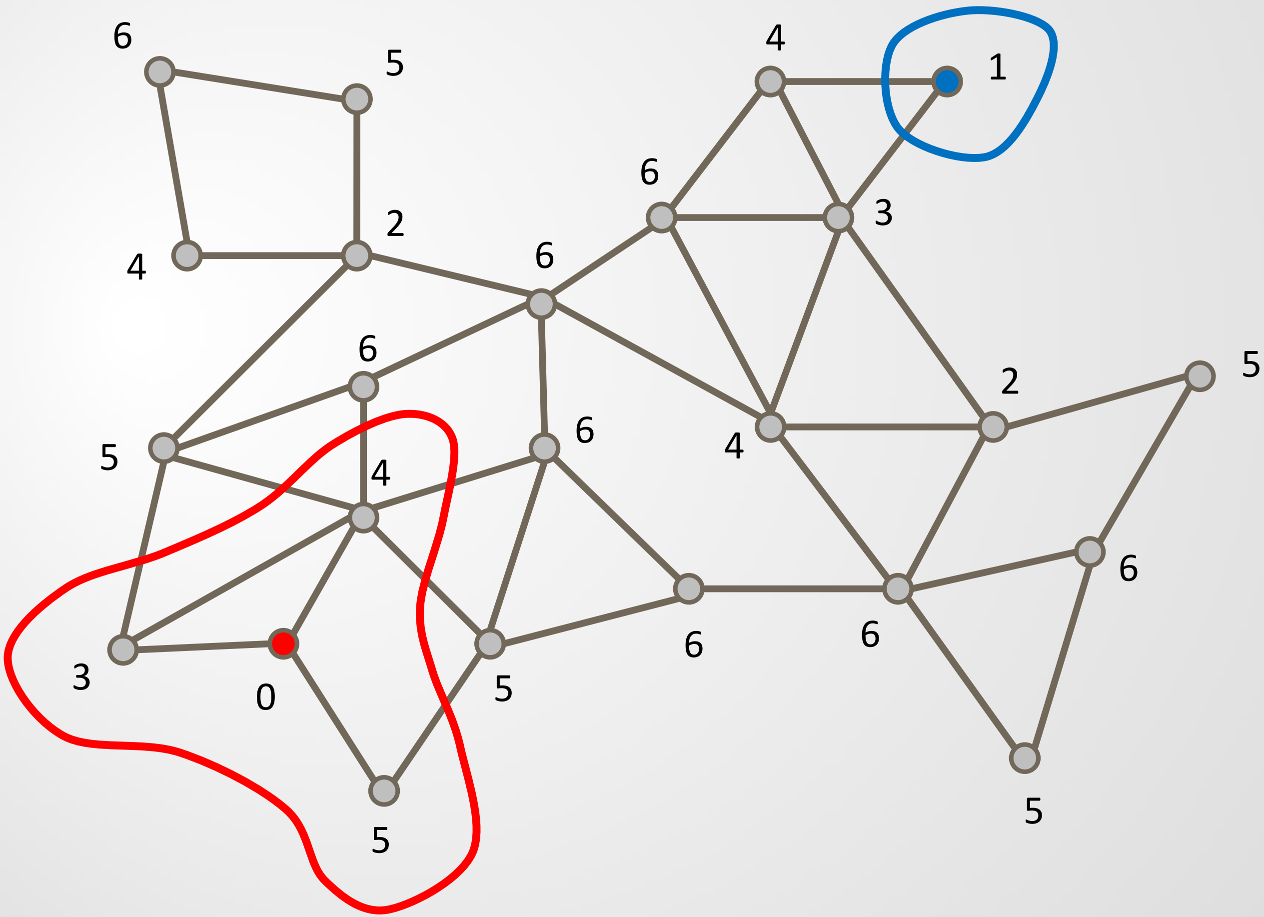
$t = 0$



F

# An Example Run [MPX13]

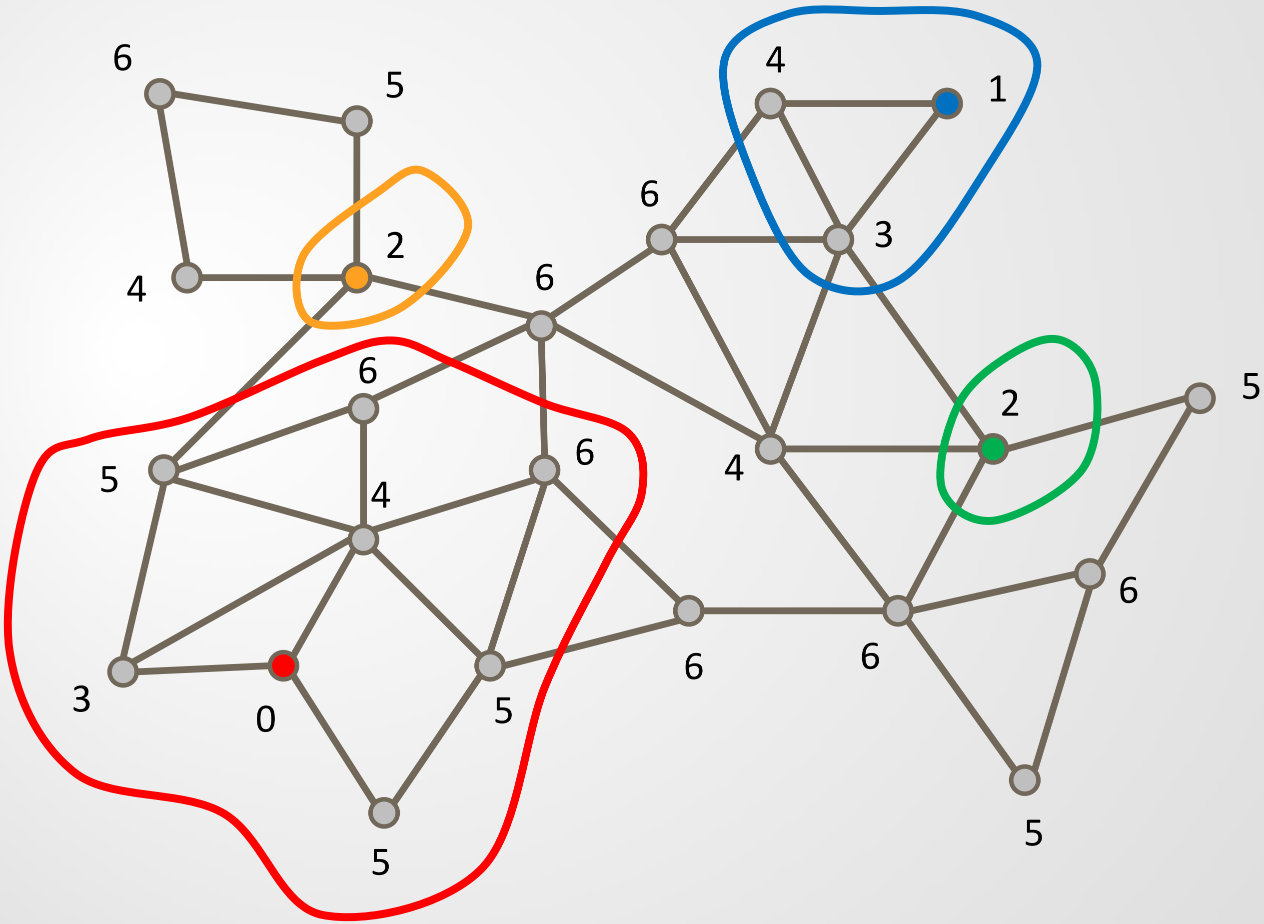
$t = 1$



G

# An Example Run [MPX13]

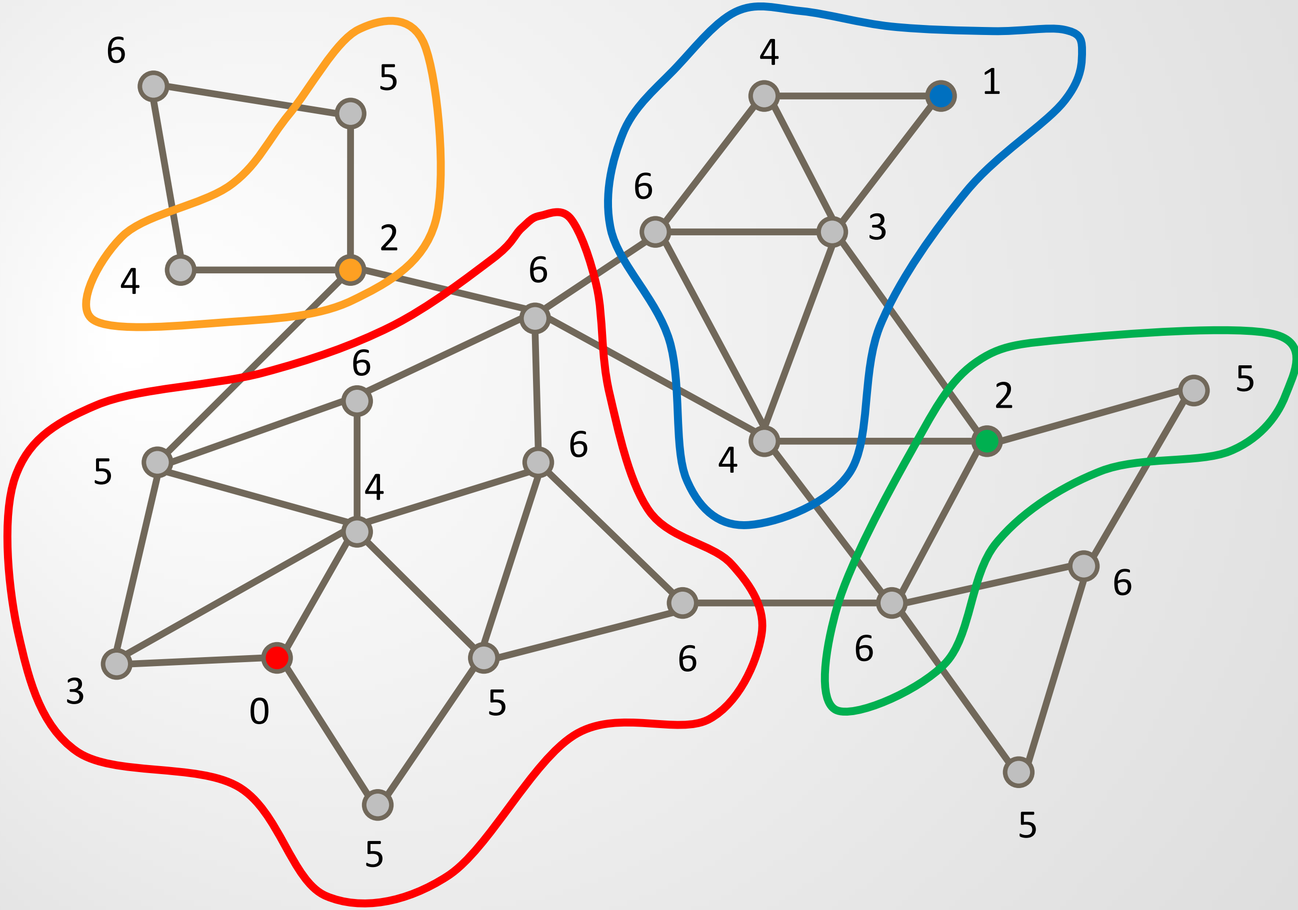
$t = 2$



H

# An Example Run [MPX13]

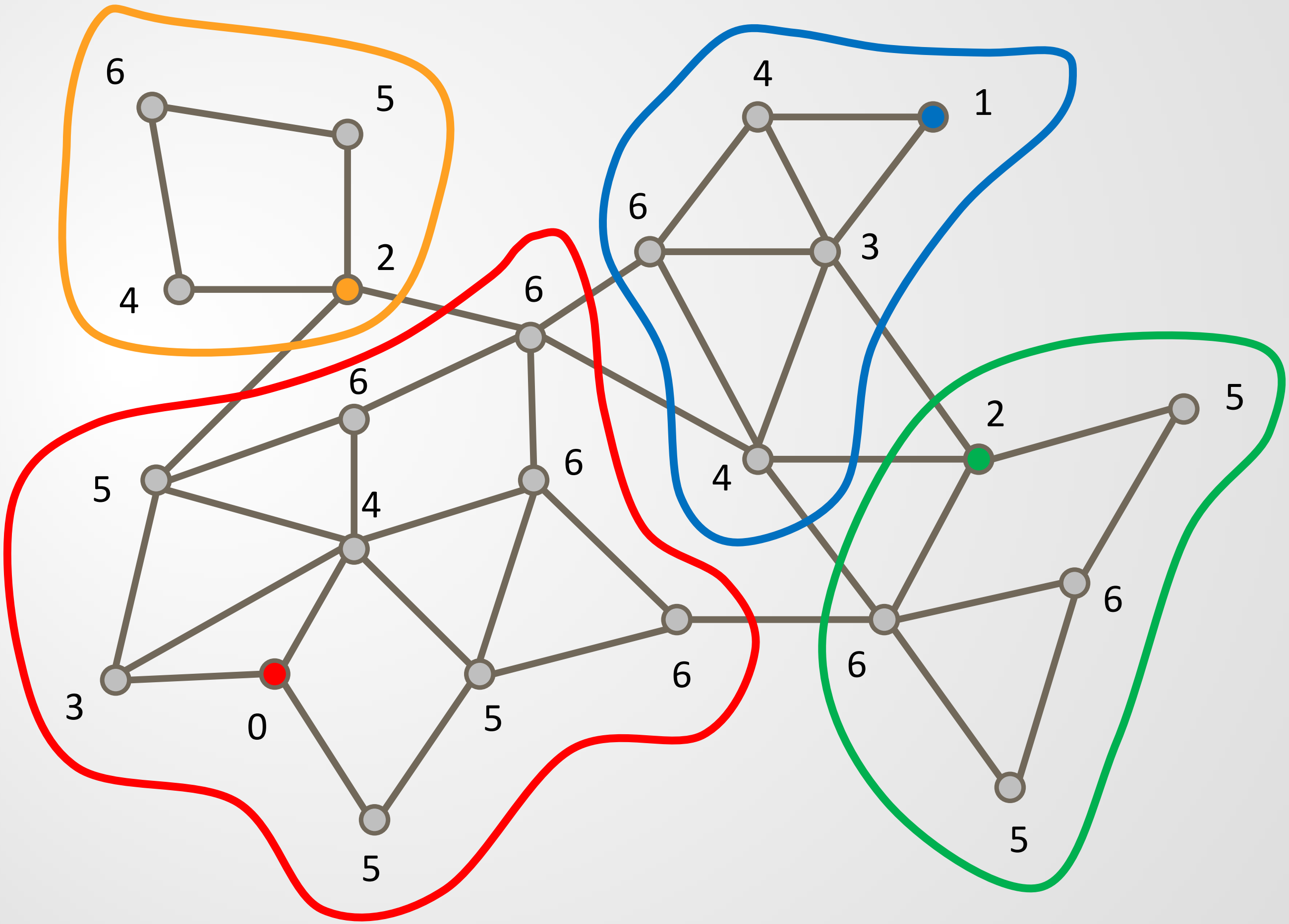
$t = 3$



I

# An Example Run [MPX13]

$t = 4$



# What is maximum cluster radius? 

Note: At time $X_{max}$ all nodes are owned!

Thus max cluster radius $\leq X_{max}$

$$X_{max} \leq \frac{2\ln n}{\beta} \text{ with prob} \geq 1 - \frac{1}{n}.$$

Question: What is prob an edge is intercluster?
i.e. prob an edge is cut!

Let $e = (u,v)$ be some edge and $c$ its midpoint.

We think of each vertex doing a BFS of $G$.
starting at time $S_{v_i} = S_i$

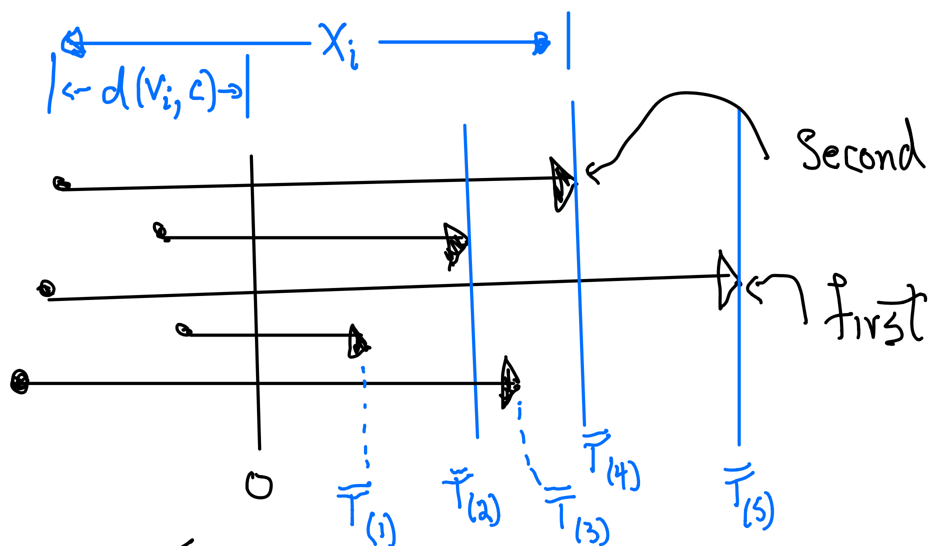
The arrival time at $c$ will be a
random variable =

$$T_i = X_{max} - X_i + dist(v_i, c)$$

Def: $\bar{T}_i = X_{max} - T_i = X_i - dist(v_i, c)$
(early arrival) (Owner has max $\bar{T}_i$)

# Horse Race & Photo Finish.

We think of $dist(V_i, C)$ as handy-cap and $X_i$ speed.




<u>Def</u> $\bar{T}_{(i)} \equiv$ ith order statistics of $\{\bar{T}_1, \dots, \bar{T}_n\}$

<u>Note</u> $\bar{T}_j = \bar{T}_{(n)}$ then $V_j$ "own" $c$ and either $u$ or $v$.

<u>Def</u> $GAP_1 = \bar{T}_{(n)} - \bar{T}_{(n-1)}$

<u>Note</u> If $GAP_1 > 1$ then $V_j$ will own $U$ & $V$

Thus $e = (u, v)$ is not "cut".

By memory less property $Gap_1 \approx Exp(\beta)$

Thus $Prob[Gap_1 < 1] = 1 - e^{-\beta}$

Claim: $1 - e^{-\beta} < \beta$

$$e^{-\beta} = 1 - \beta + \beta^2/2! - \beta^3/3!$$

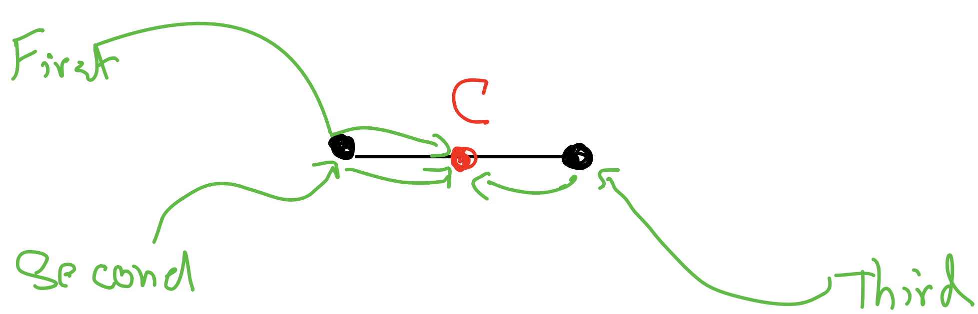
$$1 - e^{-\beta} = \beta - \beta^2/2! + \cdots < \beta \quad \text{(Taylor's Thm)}$$

Thm $\forall$ edges $e$ prob $e$ is cut or inter cluster $< \beta$.

Note: true for all edges!

Some edges may have even lower prob.

Why?



# Exponential Delay

**Thm** ExpDelay generates a clustering

1) Max Radius $\leq \ln n/\beta$ expected.
2) Max radius $\leq 2\ln n/\beta$ with prob $1-1/n$.
3) Expected number of inter-cluster edges is at most $\beta m$.
4) Run time is $O(m+n)$ Seq   $O(\log n)$ Parallel
5) (Strong Diameter Prop) If $w \in \text{Cluster}_v$ then shortest path from $v$ to $w$ is in $\text{Cluster}_v$.

---

There still maybe a large number of inter-cluster edge!

eg $m \simeq n^2$ & $\beta = 1/10 \Rightarrow n^2/10$ edge!

Will show that they are nice!

Question: How many clusters will a vertex see (share an edge with).

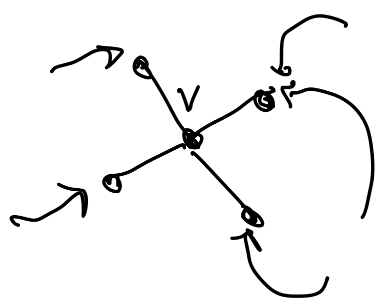
1) It will belong to one cluster.

2) How many edges to distinct clusters.

---

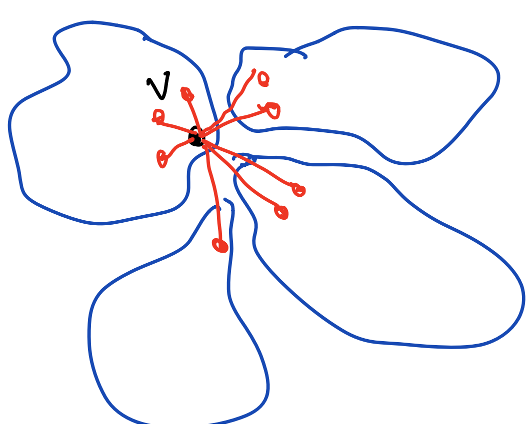
Back to horse racing.

Consider early arrivals to $v$.



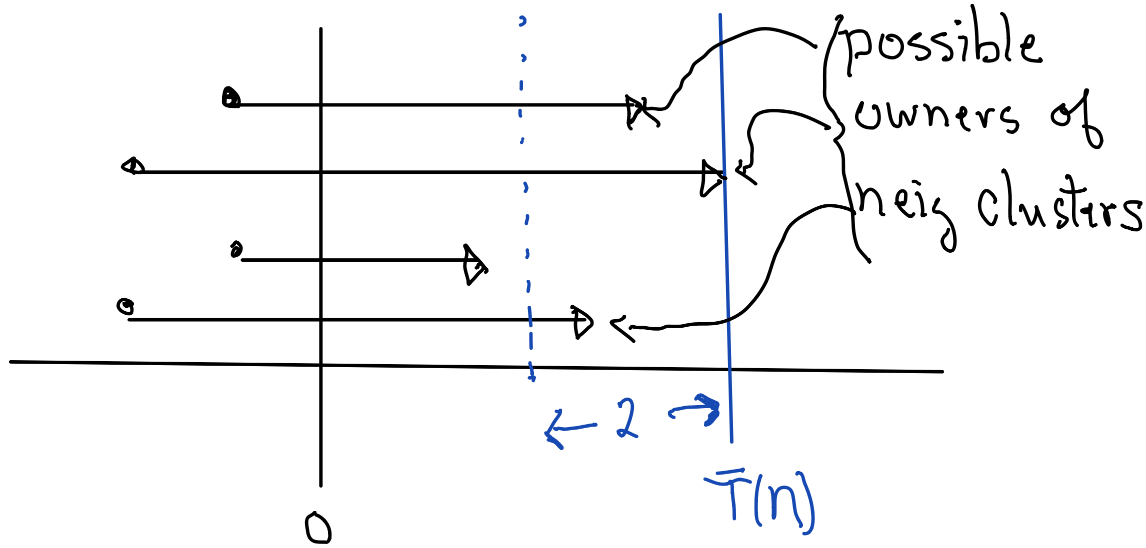

Must arrive within 2 units to possibly own a neighbor of $v$.

eg




$v$ sees 4 clusters

# Possibe Neighbor Clusters to v.




We prove a more general thm:

Suppose B is a ball of G with
1) center v.
2) diameter d.

Consider random variable

$C_B$ = Cluster(B) = $|\{\text{cluster} \mid \text{cluster} \cap B \neq \emptyset\}|$

<u>Thm</u> $\mathrm{Exp}[C_B] \leq e^{d\beta}$

$A_B$ = number of arrivals within d time of first.

Note $C_B \leq A_B$

Claim: $\mathrm{Prob}[A_B \geq t] = (1 - e^{-d\beta})^{(t-1)}$

<u>pf of claim</u>

Consider time of tth early arrival

i.e. $\overline{T}_{(n-t+1)} = T_t$

Thus at time $T_t$ there are $t-1$ memoryless exponentials still alive.

But prob of this is $(1-e^{-d\beta})^{(t-1)}$

$\square$

$$\mathbb{E}[A_B] = \sum_{t=0}^{\infty} \mathrm{Prob}[A_B \geq t] = \sum_{t=1}^{\infty} (1-e^{-d\beta})^{t-1}$$

$$= \frac{1}{1-(1-e^{-d\beta})} = e^{d\beta}$$

<u>QED</u>

$\boxed{\text{We use fact that } \sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}}$