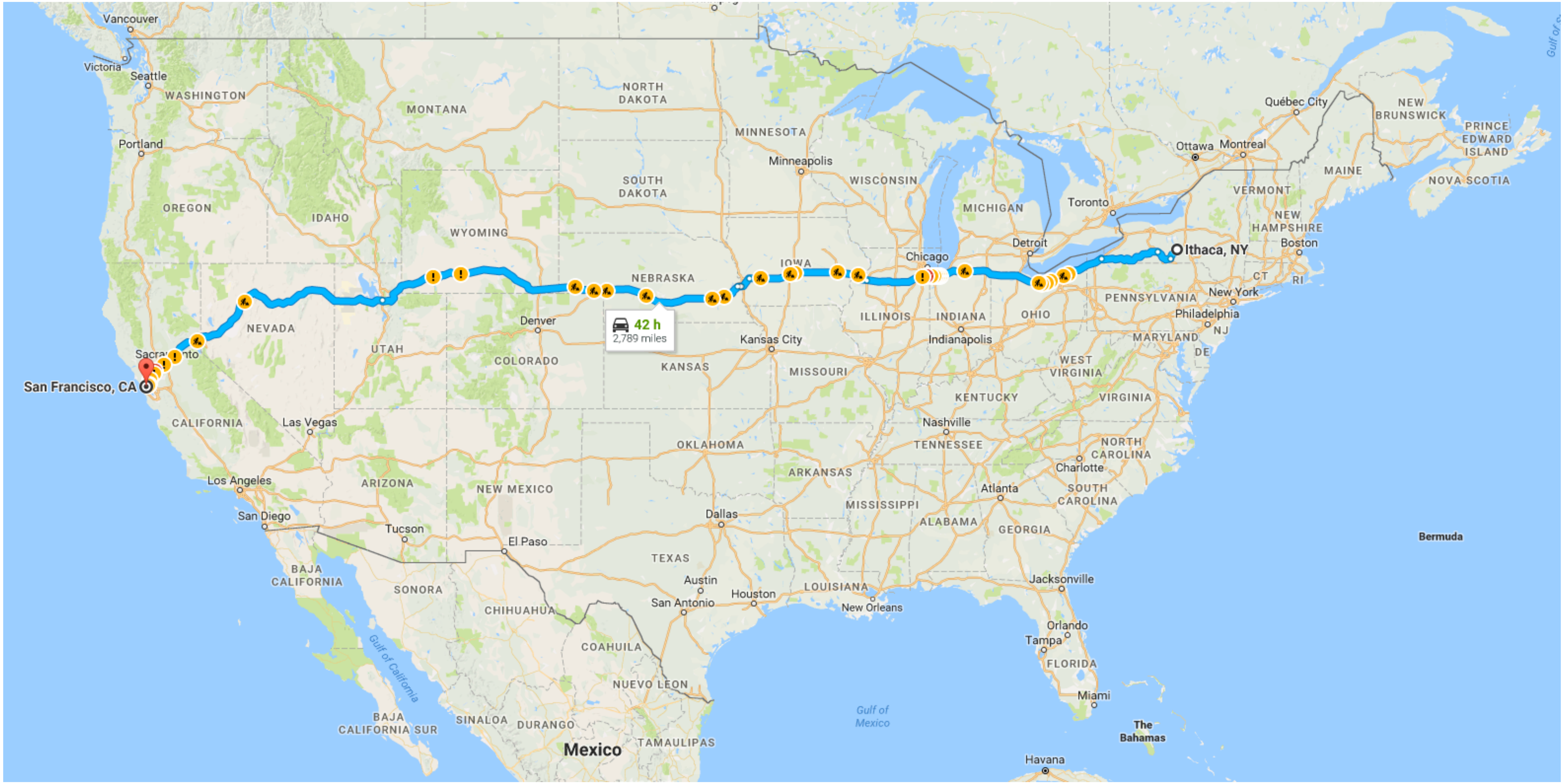
Info 2950, Lecture 17

30 Mar 2017

Prob Set 5: due Fri night 31 Mar

maps.google.com



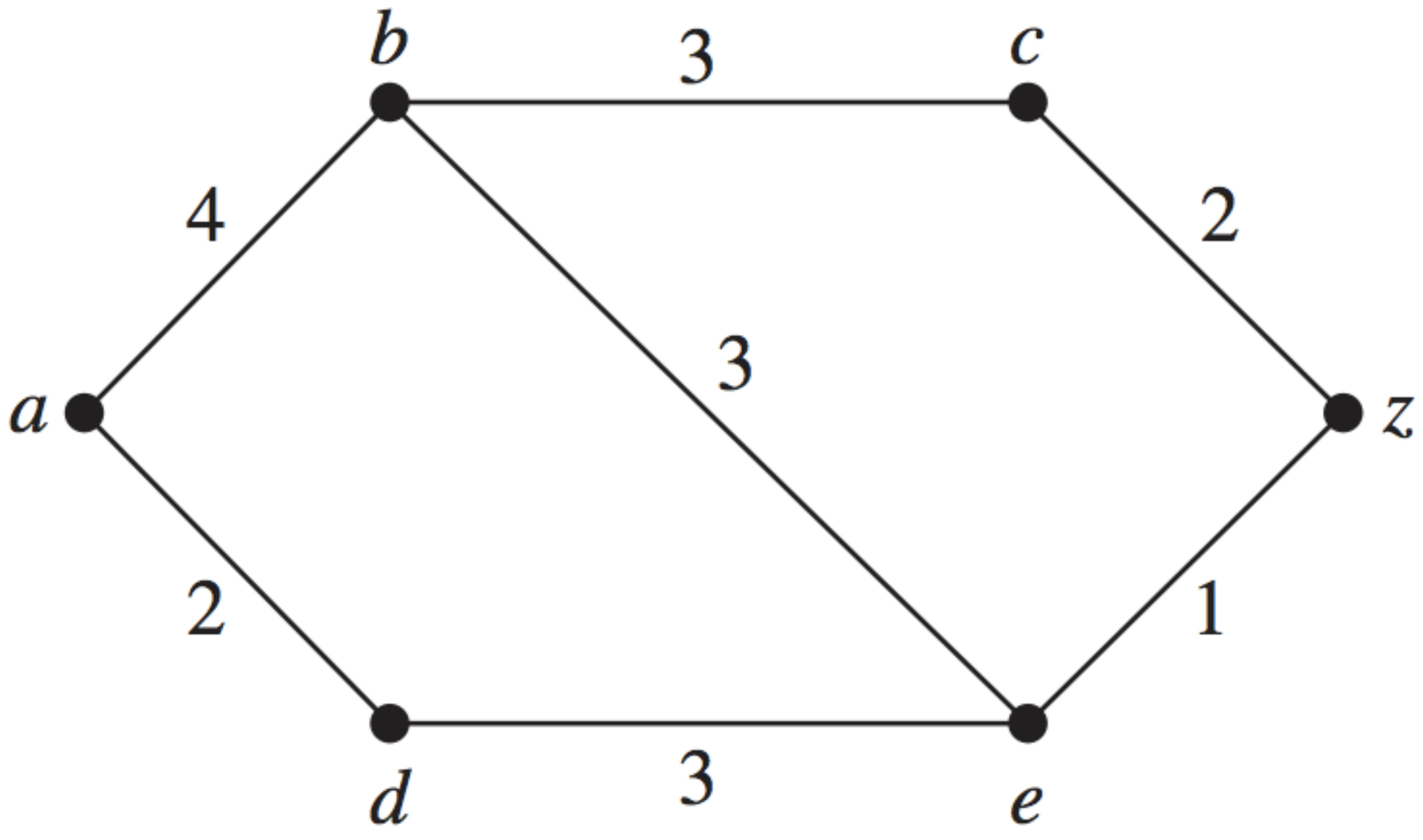
Shortest Path Algorithm

Greedy algorithm due to Dijkstra (1959)

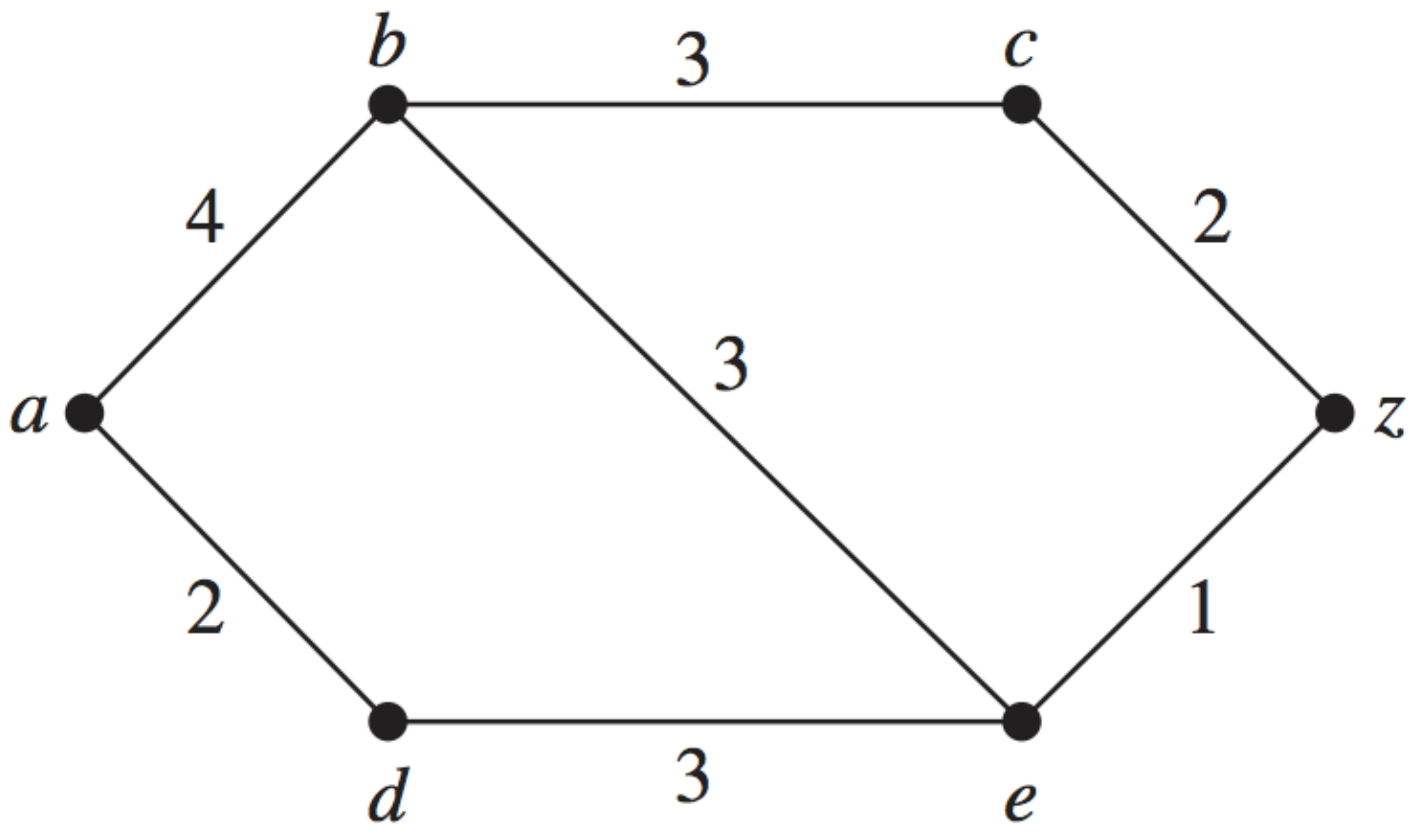
shortest path from a to z?

start from vertex a,
    find nearest vertex.
from one additional edge paths
        from those two,
    find second nearest vertex to a.
from one additional edge paths
        from those three,
    find third nearest vertex to a

and so on, until reach z



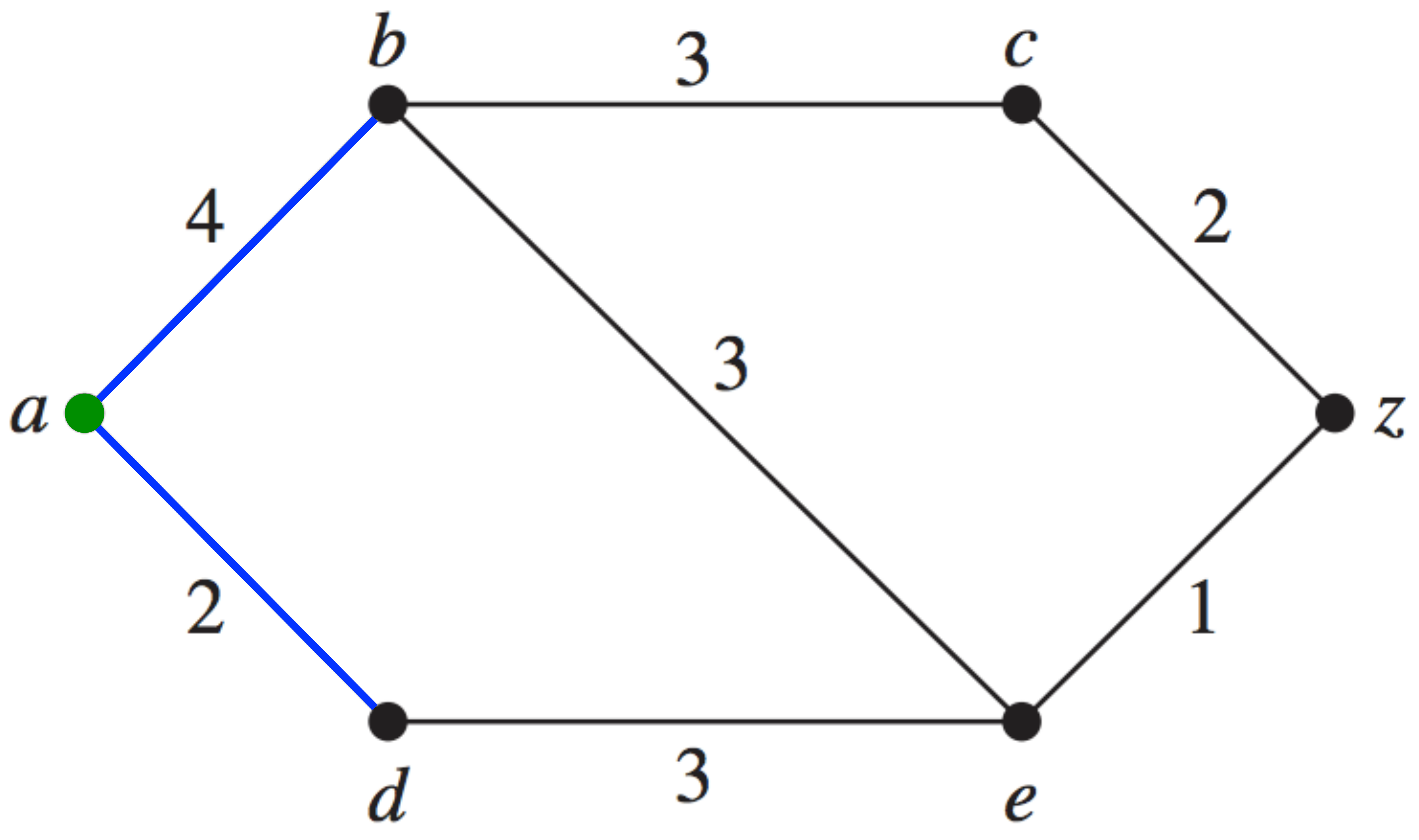
Rosen 10.6 fig.3 p.710



b
c
a
z
d
e
4
3
2
3
2
1
3

Start from {a}, two paths:

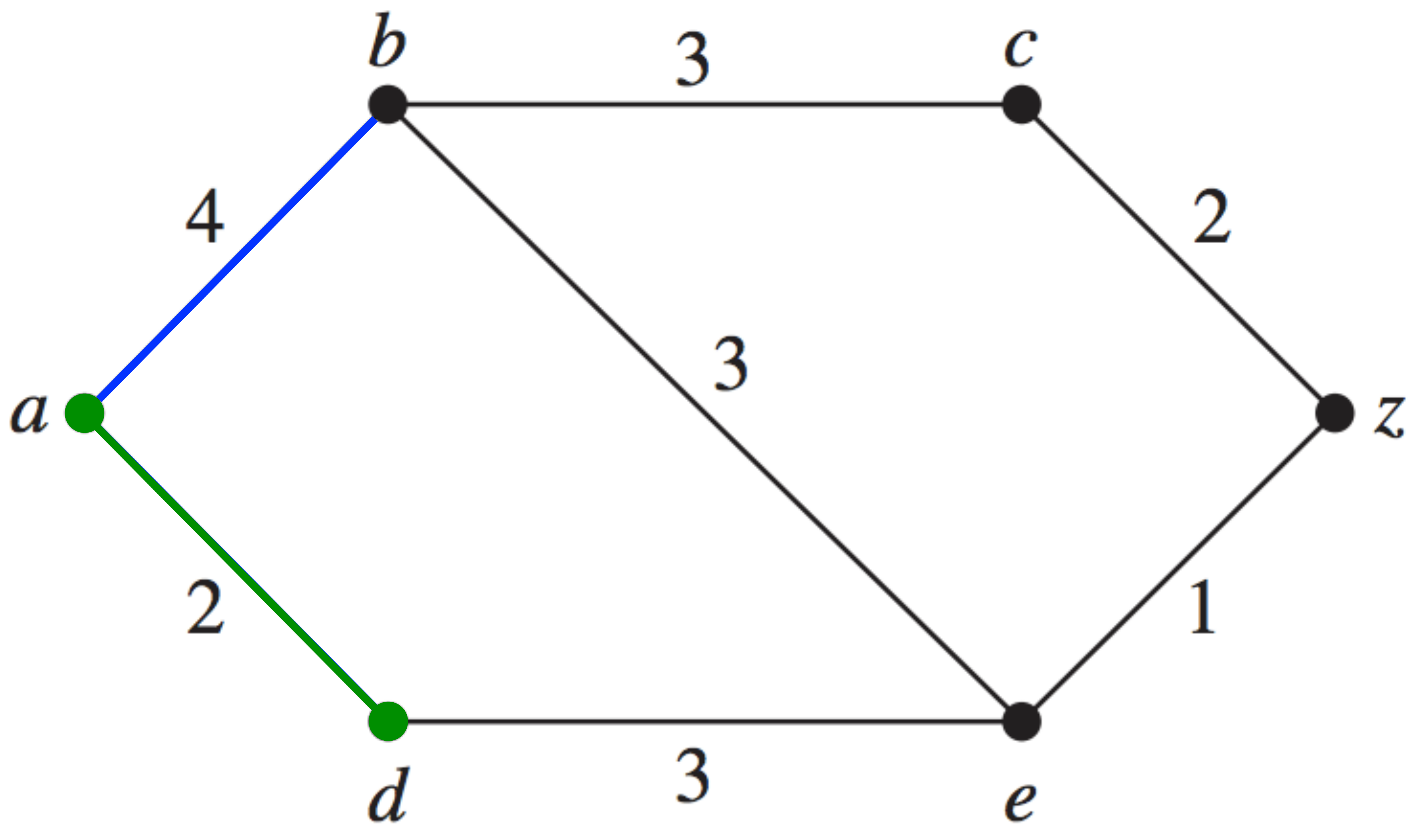
(a,b), (a,d)



Start from {a}, two paths:

(a,b), (a,d)

(a,d) is shortest



distance (a,d) = 2

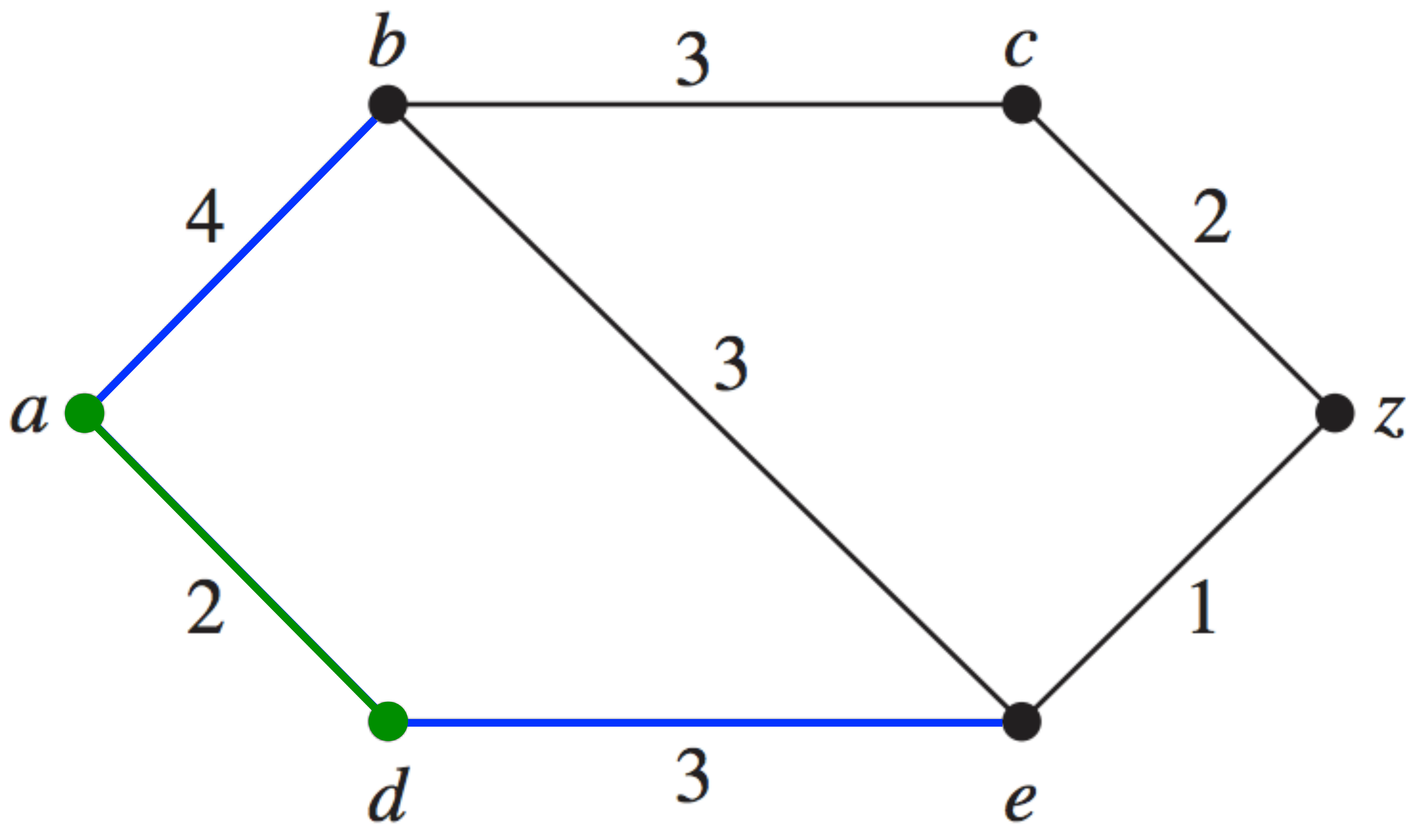
Start from {a}, two paths:

(a,b), (a,d)

(a,d) is shortest

From {a,d}, two paths:

(a,b), (a,d,e)



distance (a,d) = 2

Start from {a}, two paths:

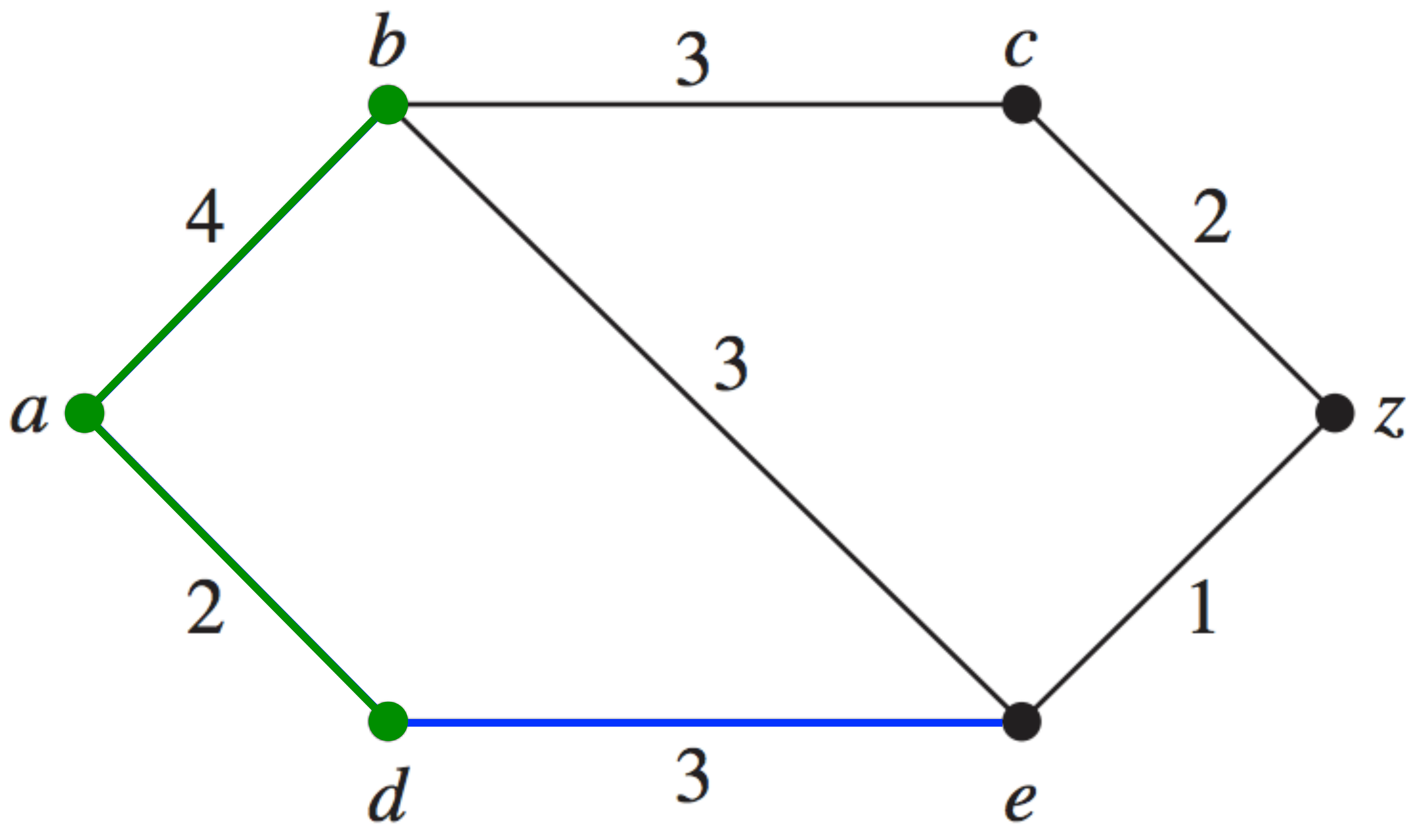
(a,b), (a,d)

(a,d) is shortest

From {a,d}, two paths:

(a,b), (a,d,e)

(a,b) is shortest



distance (a,d) = 2

distance (a,b) = 4

Start from {a}, two paths:

(a,b), (a,d)

(a,d) is shortest

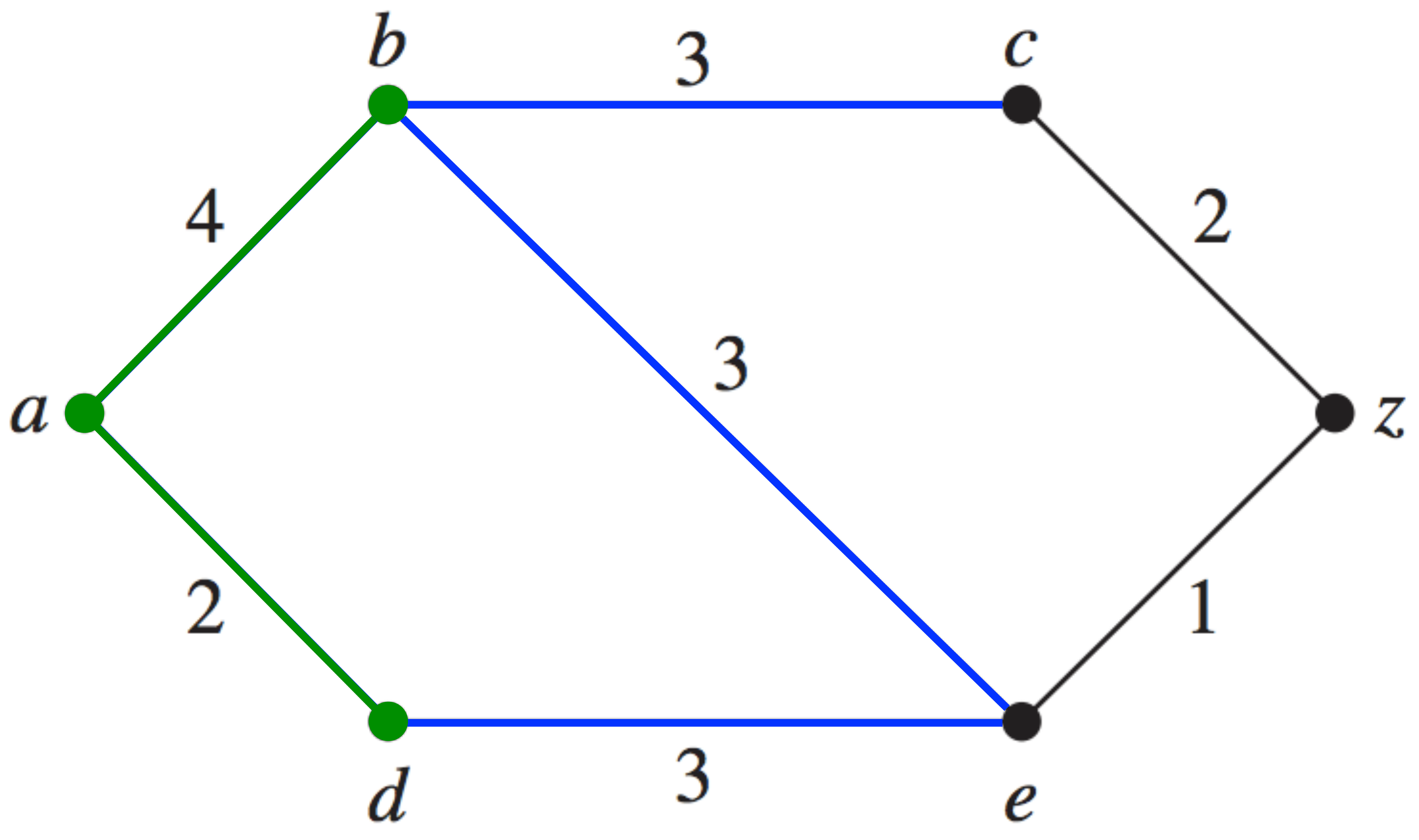
From {a,d}, two paths:

(a,b), (a,d,e)

(a,b) is shortest

From {a,b,d}, three paths:

(a,b,c), (a,b,e), (a,d,e)



distance (a,d) = 2

distance (a,b) = 4

Start from {a}, two paths:

(a,b), (a,d)

(a,d) is shortest

From {a,d}, two paths:

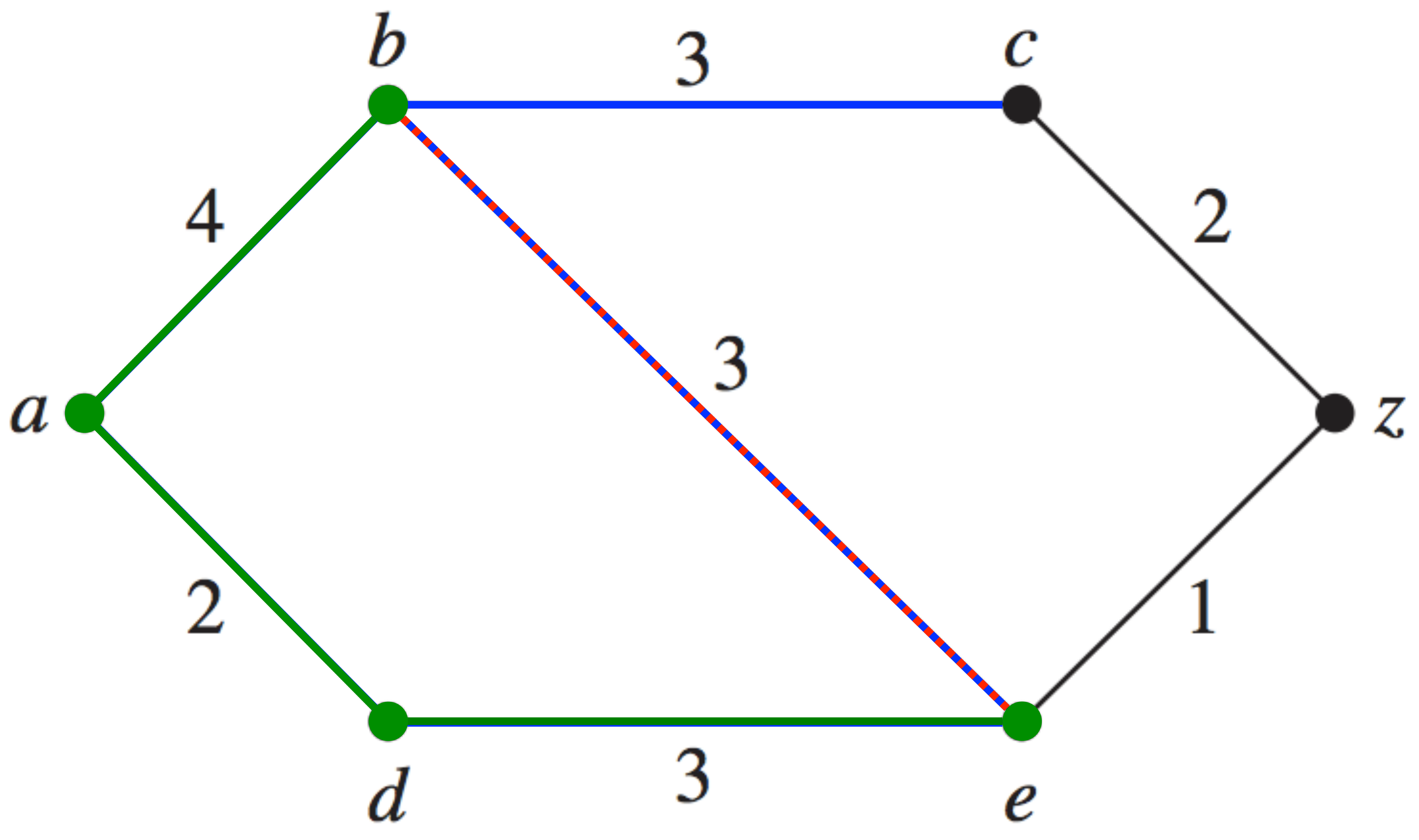
(a,b), (a,d,e)

(a,b) is shortest

From {a,b,d}, three paths:

(a,b,c), (a,b,e), (a,d,e)

(a,d,e) is shortest



distance (a,d) = 2

distance (a,b) = 4

distance (a,e) = 5

Start from {a}, two paths:

(a,b), (a,d)

(a,d) is shortest

From {a,d}, two paths:

(a,b), (a,d,e)

(a,b) is shortest

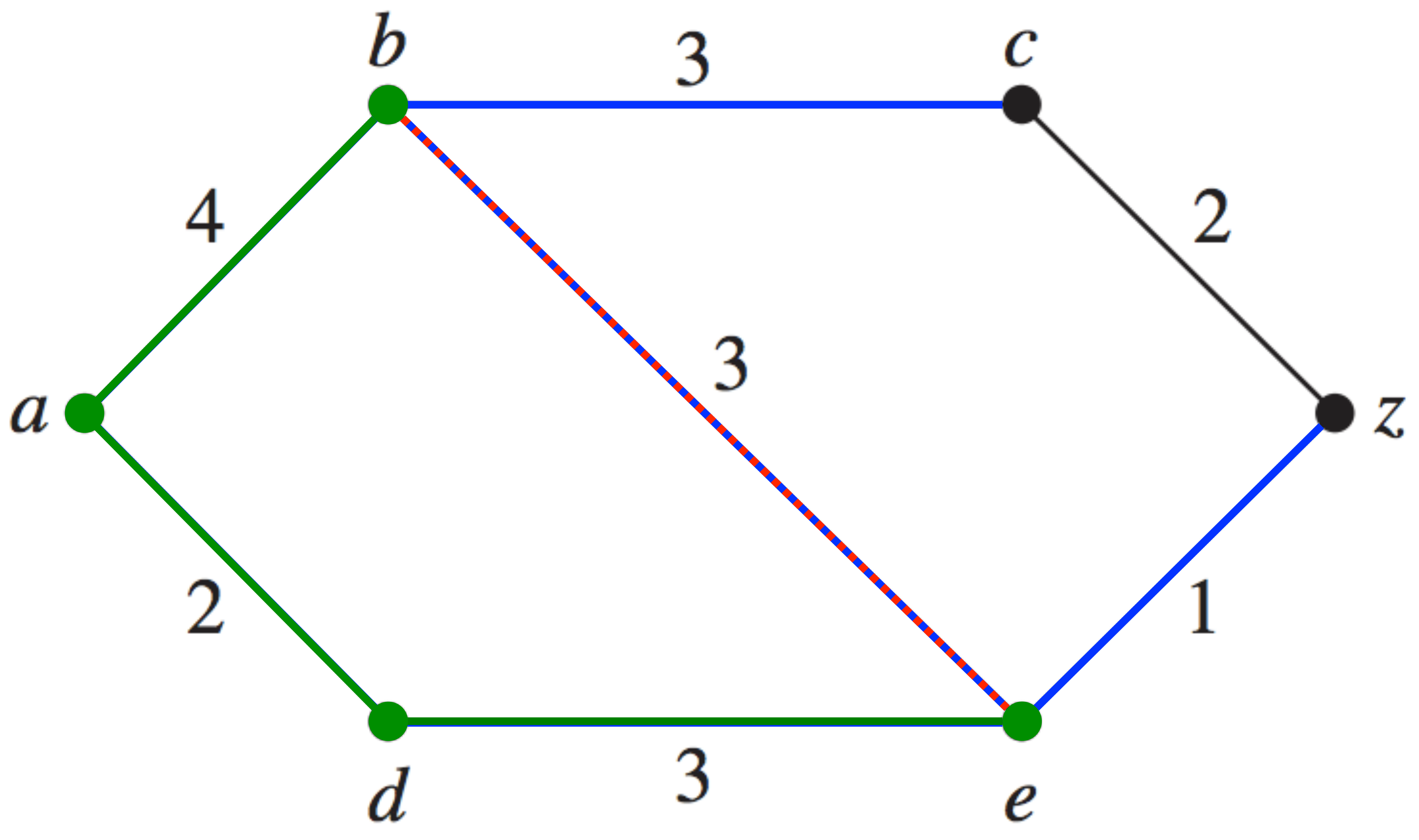
From {a,b,d}, three paths:

(a,b,c), (a,b,e), (a,d,e)

(a,d,e) is shortest

From {a,b,d,e}, two paths:

(a,b,c), (a,d,e,z)



distance (a,d) = 2

distance (a,b) = 4

distance (a,e) = 5

Start from {a}, two paths:

(a,b), (a,d)

(a,d) is shortest

From {a,d}, two paths:

(a,b), (a,d,e)

(a,b) is shortest

From {a,b,d}, three paths:

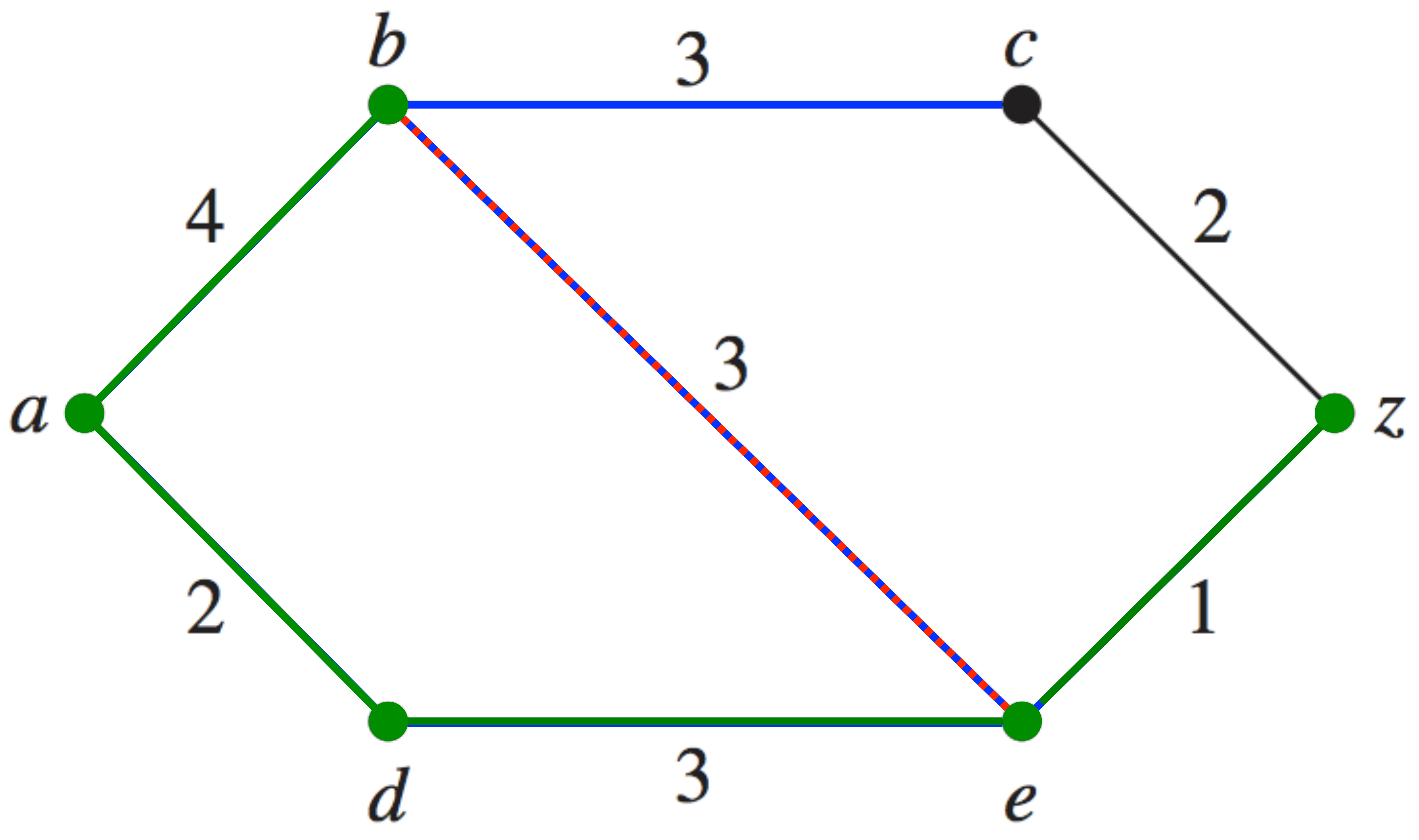
(a,b,c), (a,b,e), (a,d,e)

(a,d,e) is shortest

From {a,b,d,e}, two paths:

(a,b,c), (a,d,e,z)

(a,d,e,z) is shortest



distance (a,d) = 2

distance (a,b) = 4

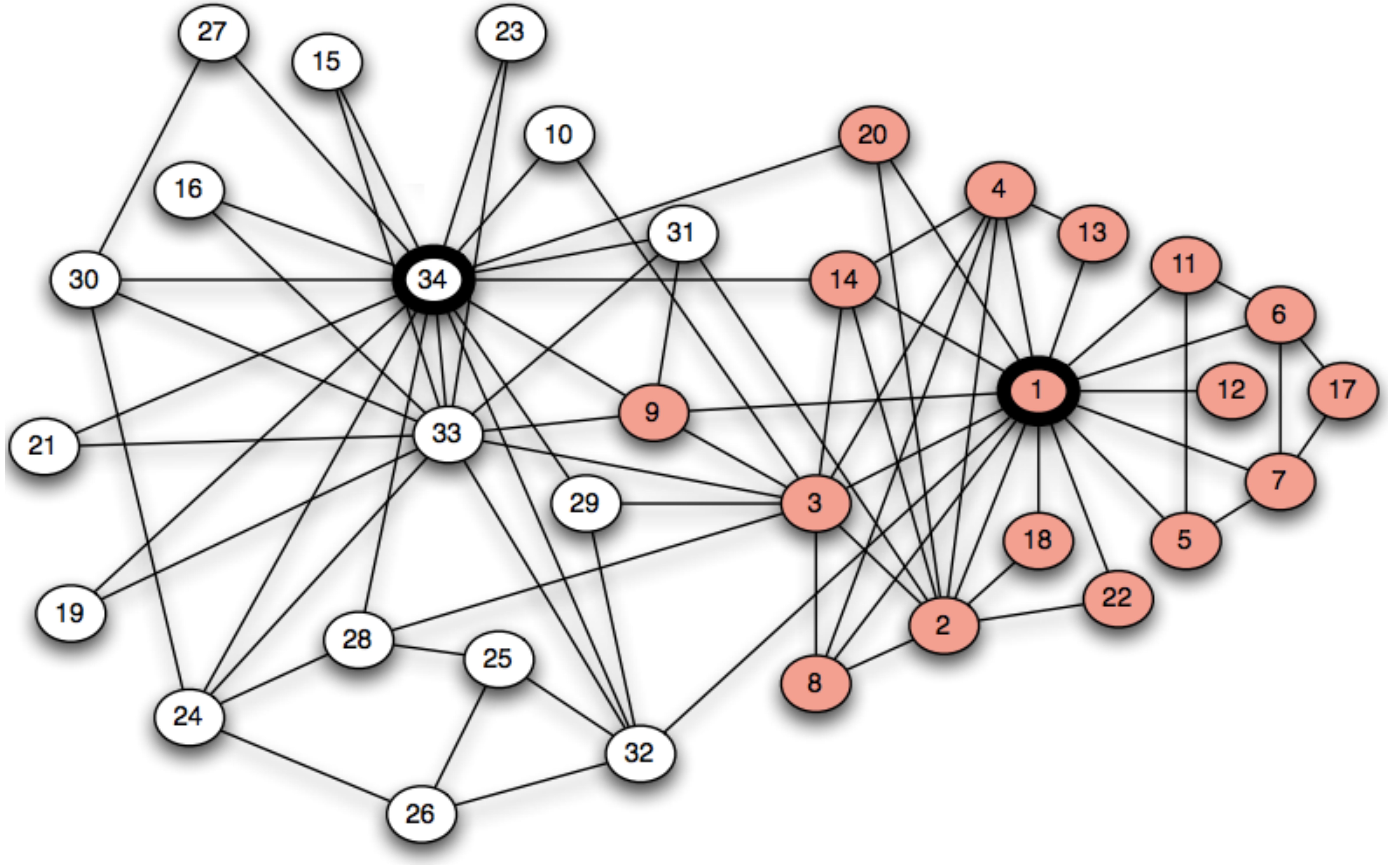
distance (a,e) = 5

distance (a,z) = 6

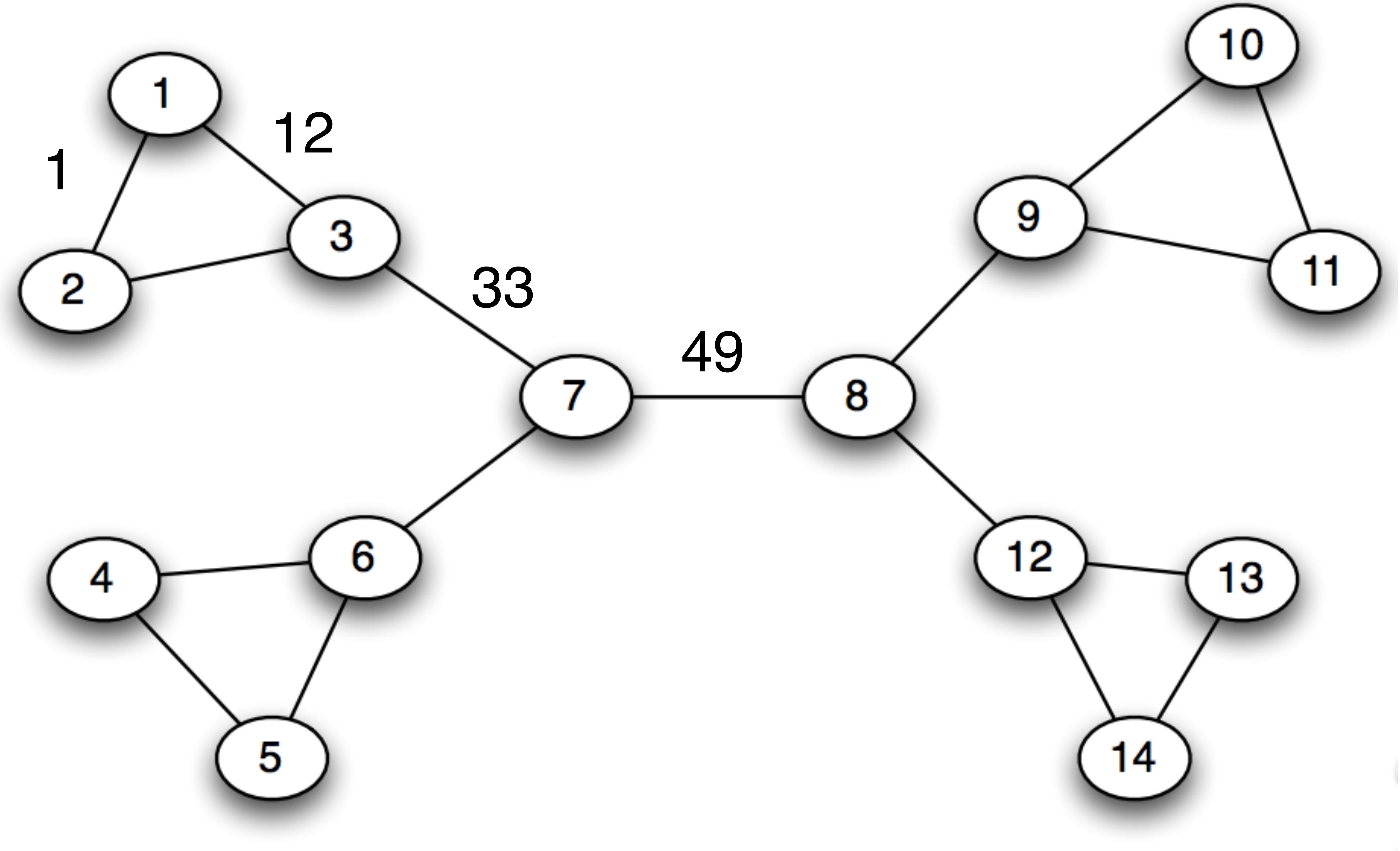
https://en.wikipedia.org/wiki/Zachary's_karate_club

“A social network of a karate club was studied by Wayne W. Zachary for a period of three years from 1970 to 1972. The network captures 34 members of a karate club, documenting 78 pairwise links between members who interacted outside the club. During the study a conflict arose between the club president [34] and instructor [1], which led to the split of the club into two. Half of the members formed a new club around the instructor, members from the other part found a new instructor or gave up karate. Based on collected data Zachary assigned correctly all but one member of the club to the groups actually joined after the split.”

Easley/Kleinberg fig 3.13: Could the boundaries of the two subclubs be predicted from the network structure?

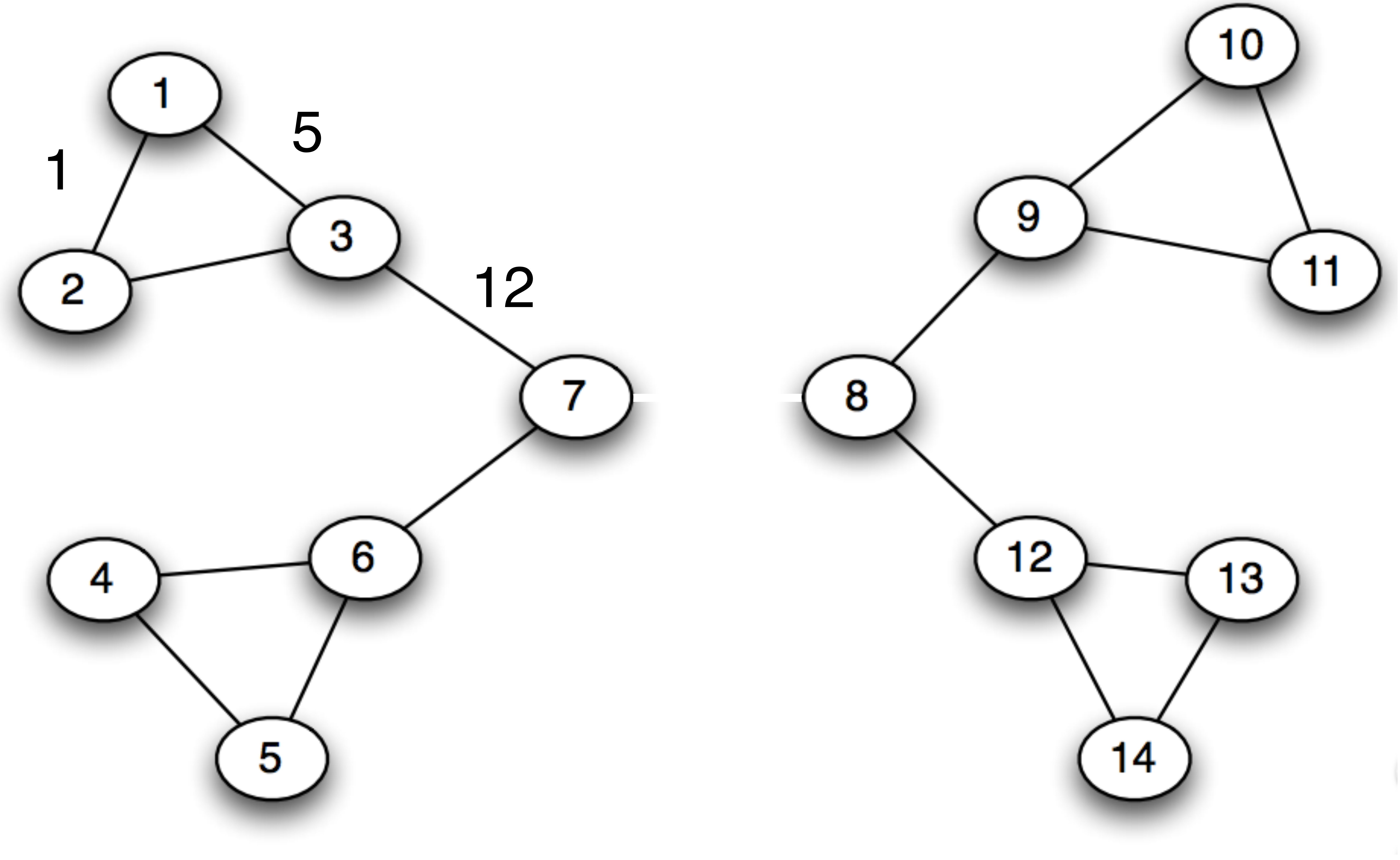


use “Edge betweenness”= # shortest length paths through edge as splitting criterion



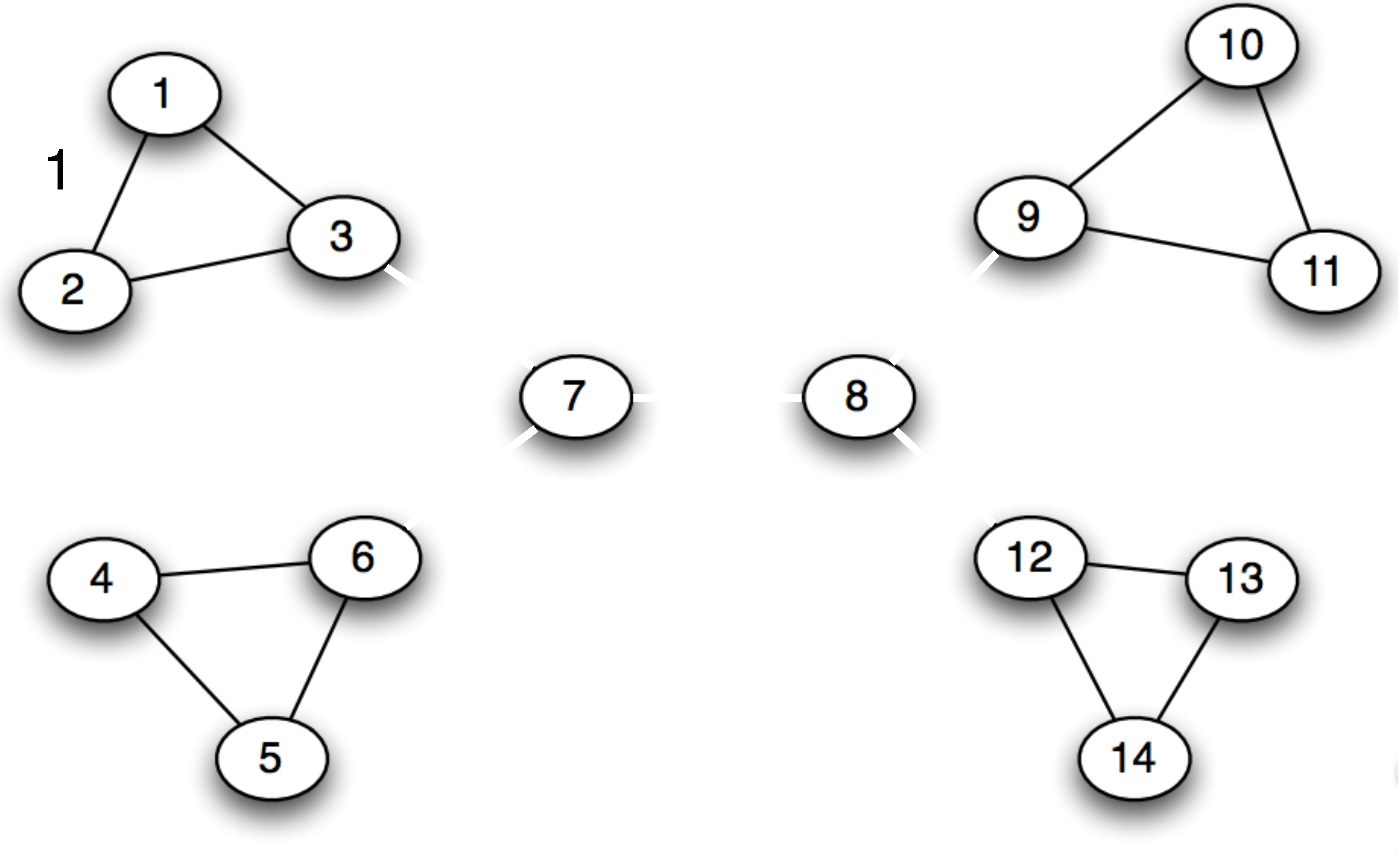
Easley/Kleinberg fig 3.14a

use “Edge betweenness”= # shortest length paths through edge as splitting criterion



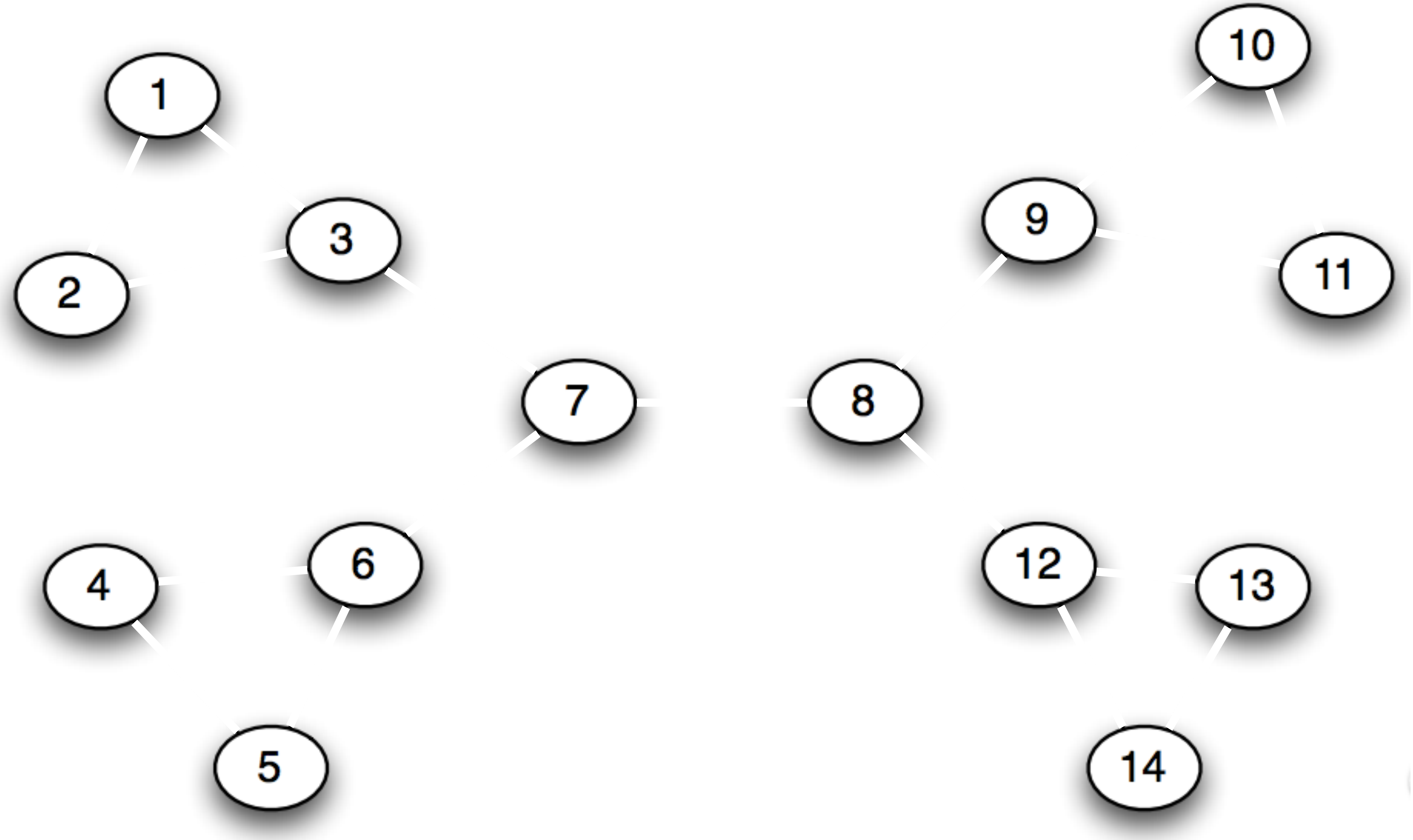
Easley/Kleinberg fig 3.14a

use “Edge betweenness”= # shortest length paths through edge as splitting criterion

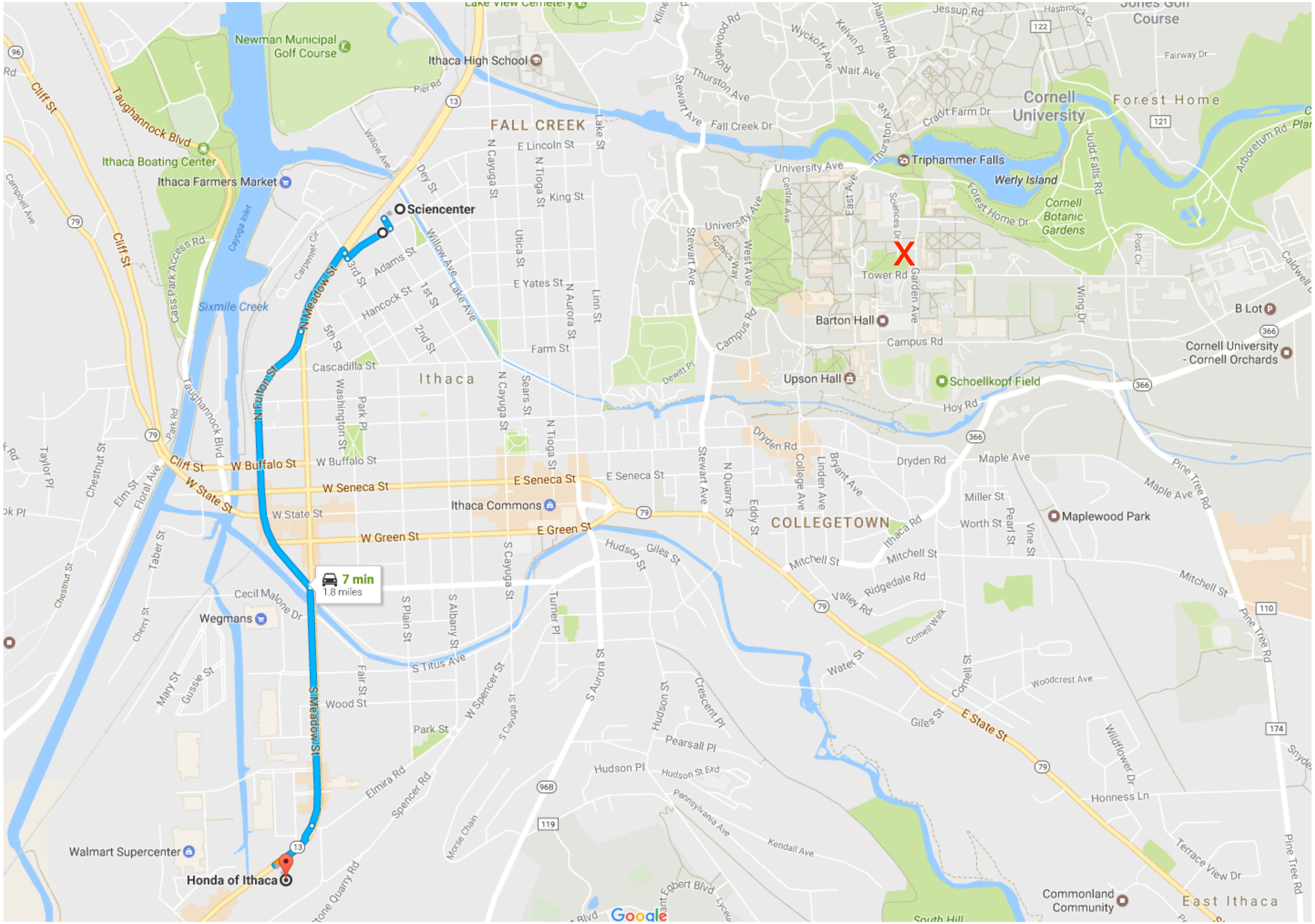


Easley/Kleinberg fig 3.14a

use “Edge betweenness”= # shortest length paths through edge as splitting criterion



Easley/Kleinberg fig 3.14a


Newman Municipal Golf Course
Ithaca High School
Ithaca Boating Center
Ithaca Farmers Market
Sciencenter
FALL CREEK
Cornell University
Forest Home
Triphammer Falls
Werly Island
Cornell Botanic Gardens
X
Tower Rd
Barton Hall
Campus Rd
Upson Hall
Schoellkopf Field
Hoy Rd
B Lot
Cornell University - Cornell Orchards
Ithaca
Ithaca Commons
COLLEGETOWN
Maplewood Park
7 min
1.8 miles
Wegmans
Walmart Supercenter
Honda of Ithaca
Commonland Community
East Ithaca
Sixmile Creek
Cayuga Inlet
Taughannock Blvd
Cliff St
W Buffalo St
W Seneca St
W State St
W Green St
E Seneca St
E Green St
E State St
N Meadow St
N Fulton St
S Meadow St
University Ave
Stewart Ave
Dryden Rd
Mitchell St
Pine Tree Rd
Google

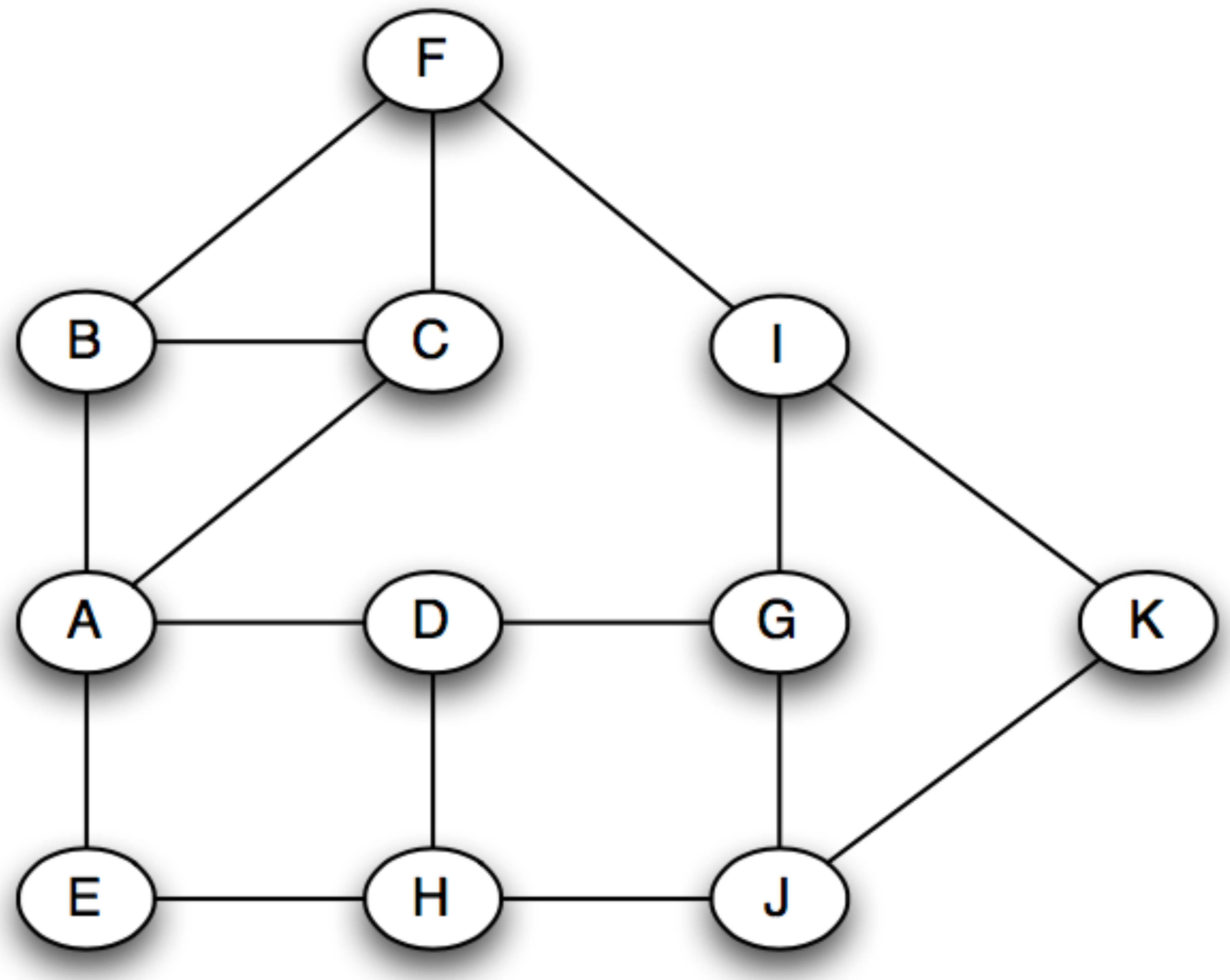
Easley/Kleinberg p.78

Which are the important vertices?
Which are the important edges?

Which play a central role
in mediating communication?

Use edge betweenness
and vertex betweenness:

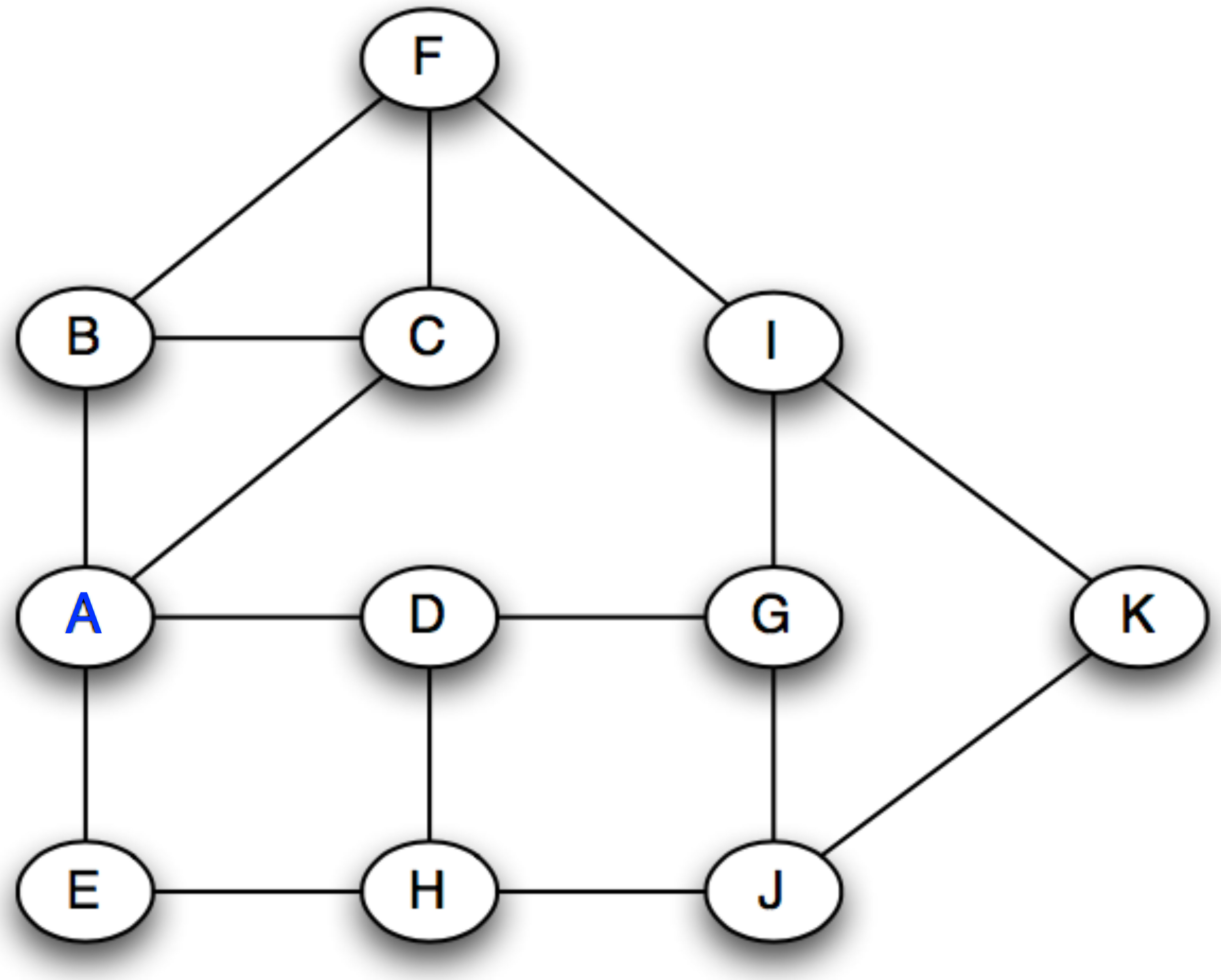
those with the most
shortest length paths going through them
may be most "central" to the network.



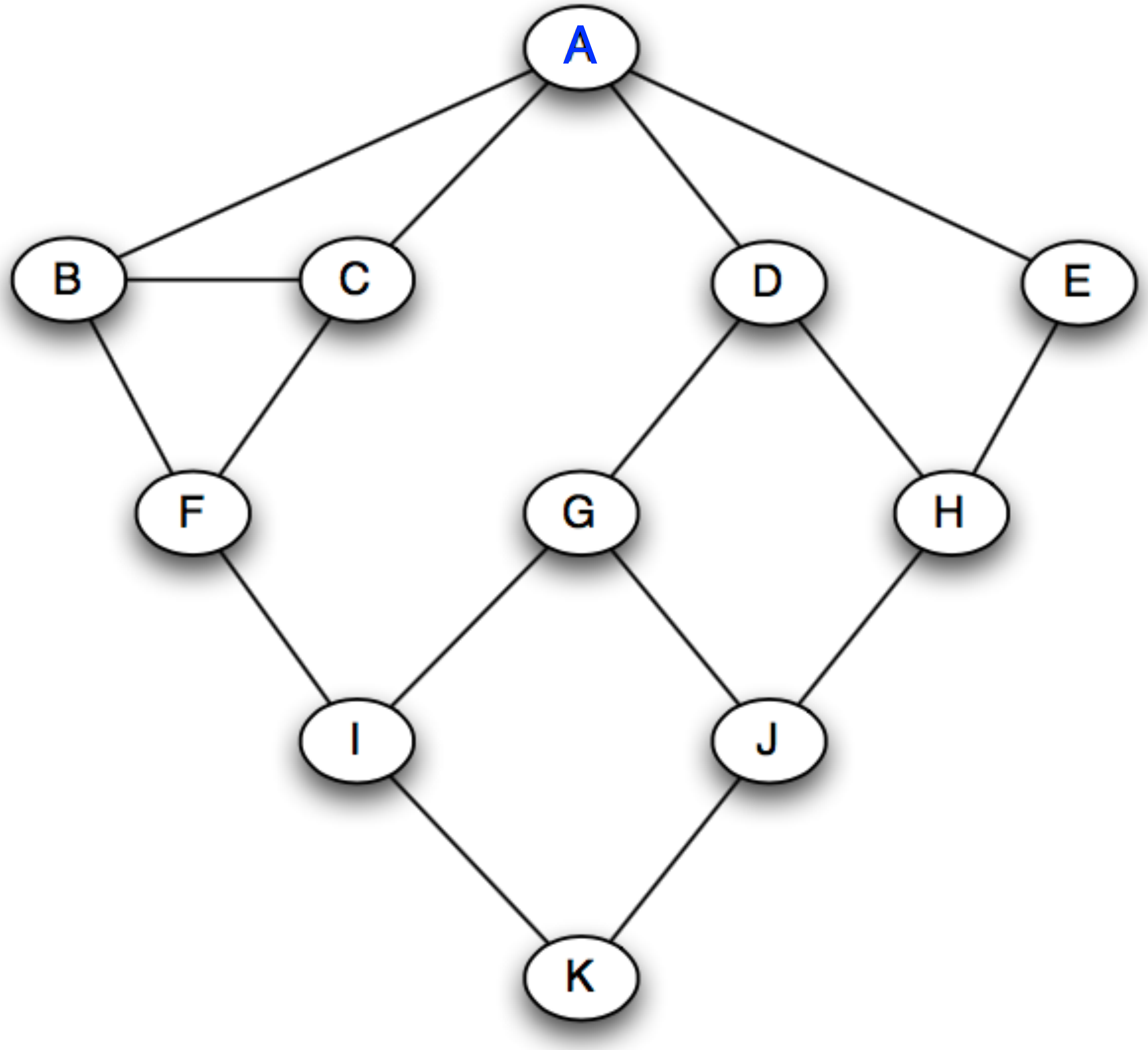
(a) *A sample network*

Easley/Kleinberg p.78

To consider centrality for paths starting from A, rearrange to put A at the top (will have to repeat for each of the nodes, so an O(N^2) algorithm)

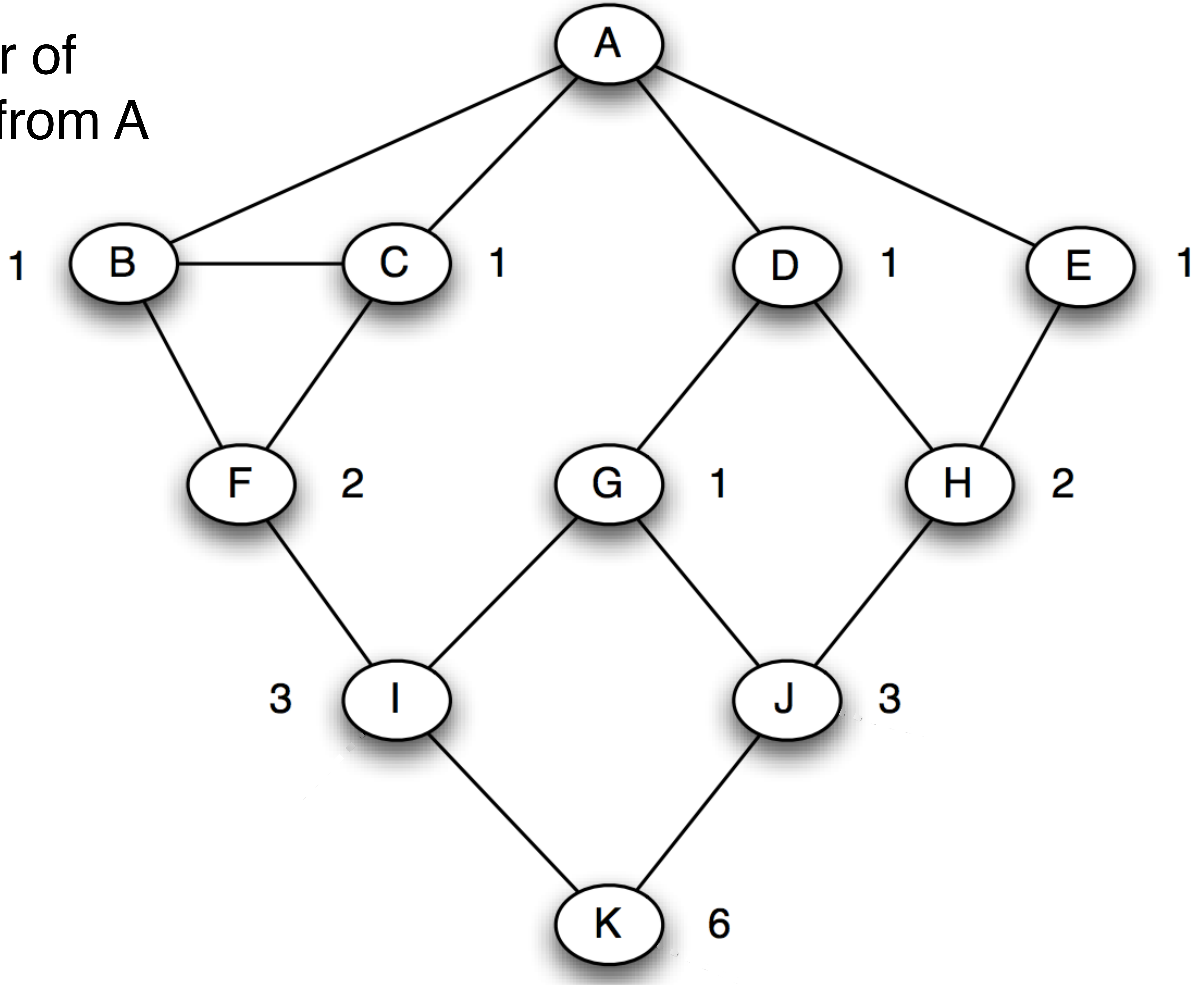


(a) *A sample network*

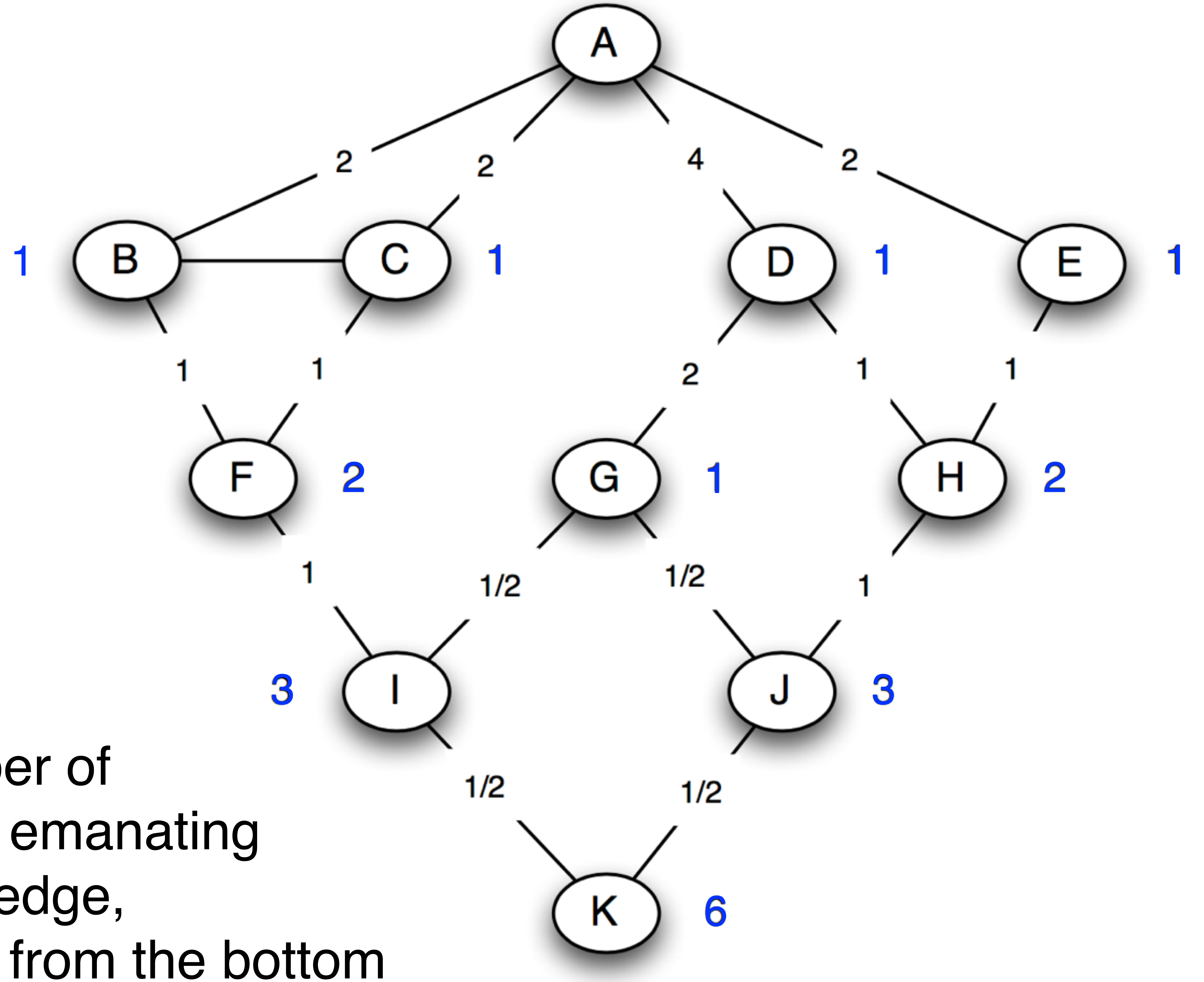


(b) *Breadth-first search starting at node A*

First count the number of shortest length paths from A to every other vertex, level by level starting from the top



(Number of shortest length paths from A to each other vertex now labelled in blue)



Then count the number of shortest length paths emanating from A through each edge, level by level starting from the bottom

Poisson Distribution

Bernoulli process with N trials, each probability p of success:

$$p(m) = \binom{N}{m} p^m (1-p)^{N-m} \, .$$

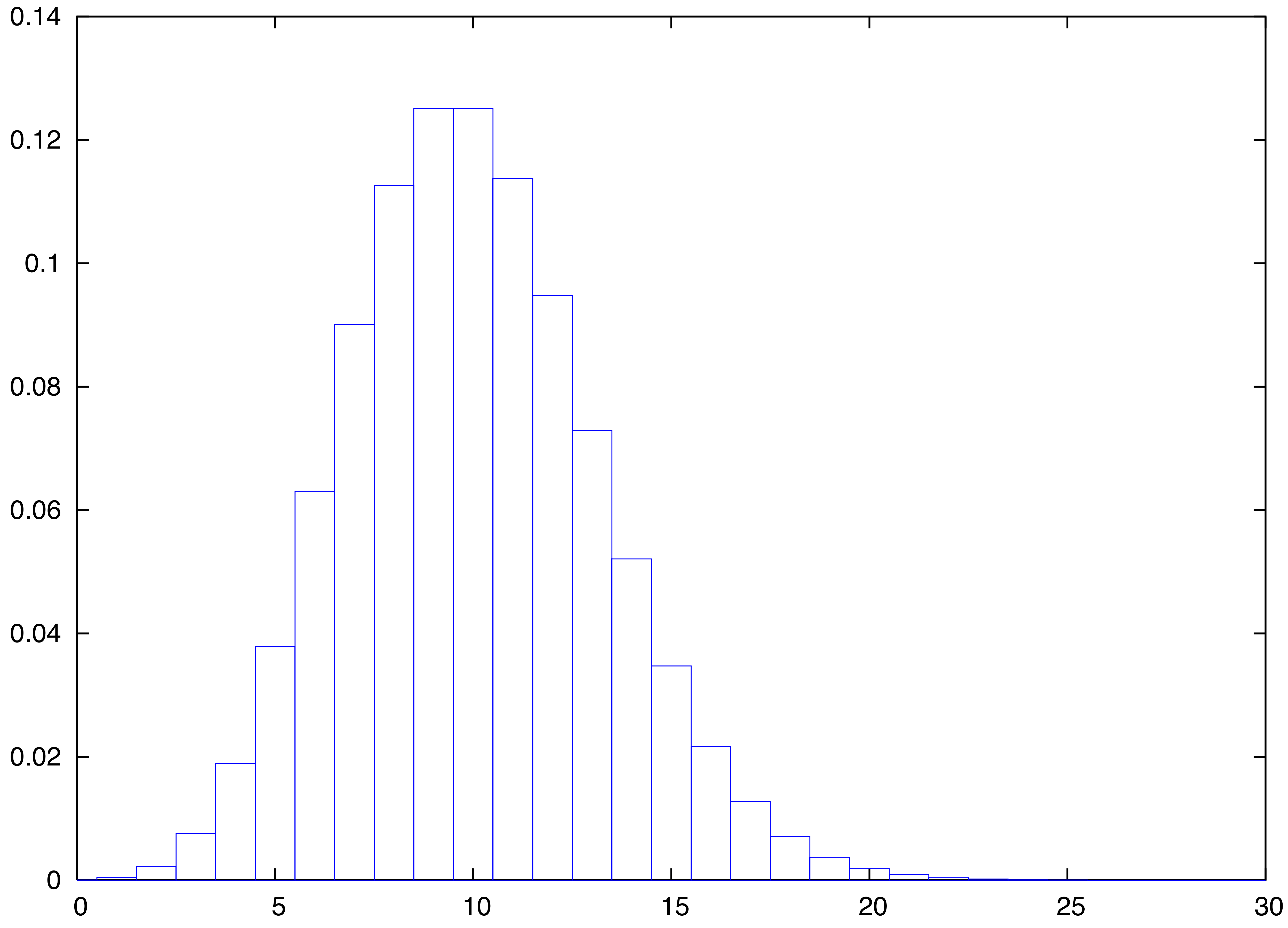
Probability $p(m)$ of m successes, in limit N very large and p small, parametrized by just $\mu = Np$ ($\mu =$ mean number of successes).
For $N \gg m$, we have $\frac{N!}{(N-m)!} = N(N-1)\cdots(N-m+1) \approx N^m$,
so $\binom{N}{m} \equiv \frac{N!}{m!(N-m)!} \approx \frac{N^m}{m!}$, and

$$p(m) \approx \frac{1}{m!} N^m \left(\frac{\mu}{N}\right)^m \left(1-\frac{\mu}{N}\right)^{N-m} \approx \frac{\mu^m}{m!} \lim_{N\to\infty} \left(1-\frac{\mu}{N}\right)^N = \mathrm{e}^{-\mu} \frac{\mu^m}{m!}$$

(ignore $(1-\mu/N)^{-m}$ since by assumption $N \gg \mu m$).
N dependence drops out for $N \to \infty$, with average μ fixed ($p \to 0$).
The form $p(m) = \mathrm{e}^{-\mu} \frac{\mu^m}{m!}$ is known as a Poisson distribution
(properly normalized: $\sum_{m=0}^{\infty} p(m) = \mathrm{e}^{-\mu} \sum_{m=0}^{\infty} \frac{\mu^m}{m!} = \mathrm{e}^{-\mu} \cdot \mathrm{e}^{\mu} = 1$).

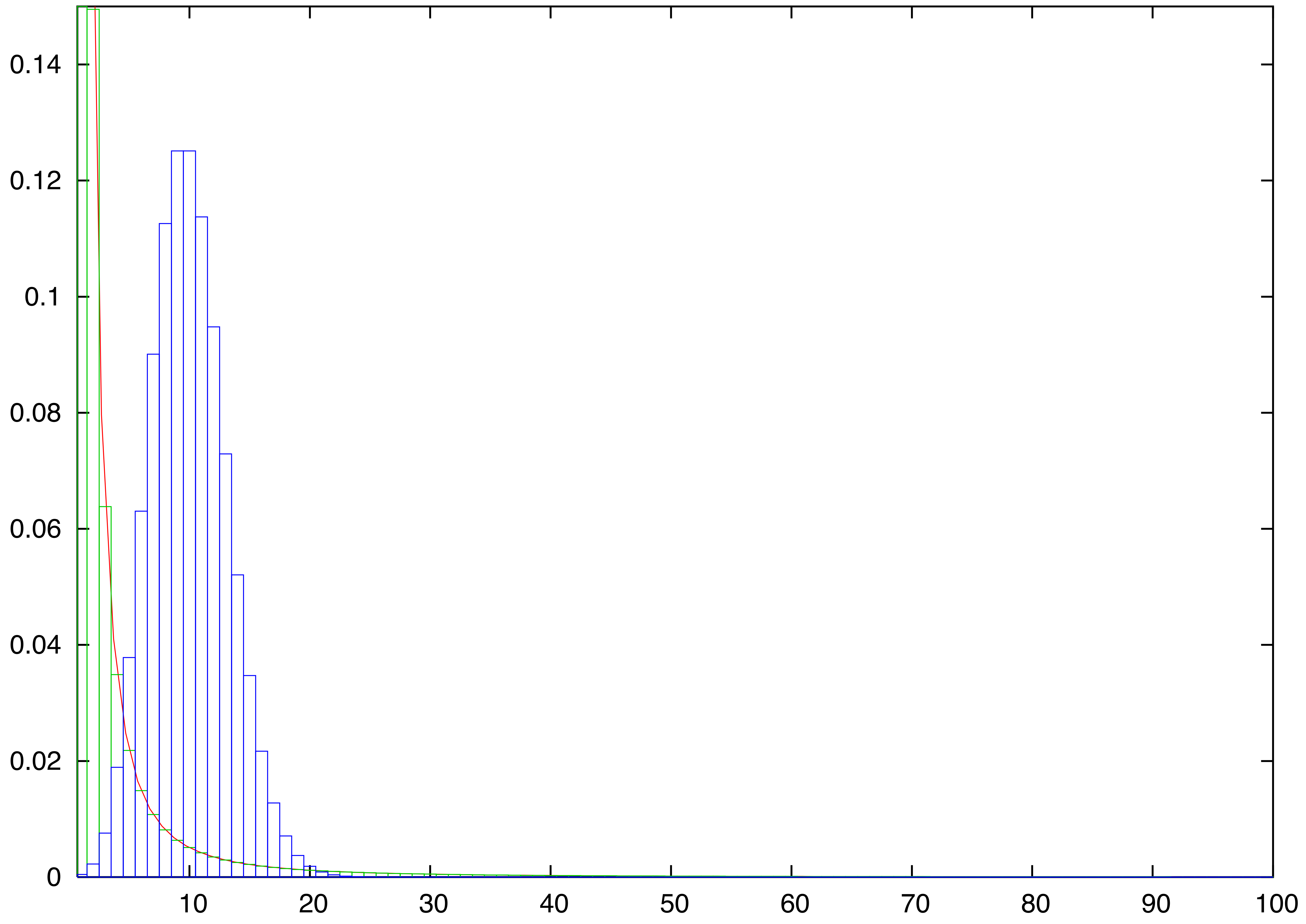
Poisson Distribution for $\mu = 10$

$$p(m) = \mathrm{e}^{-10} \frac{10^m}{m!}$$



Compare to power law $p(m) \propto 1/m^{2.1}$

Power Law $p(m) \propto 1/m^{2.1}$ and Poisson $p(m) = \mathrm{e}^{-10}\frac{10^m}{m!}$



Power Law $p(m) \propto 1/m^{2.1}$ and Poisson $p(m) = \mathrm{e}^{-10}\frac{10^m}{m!}$