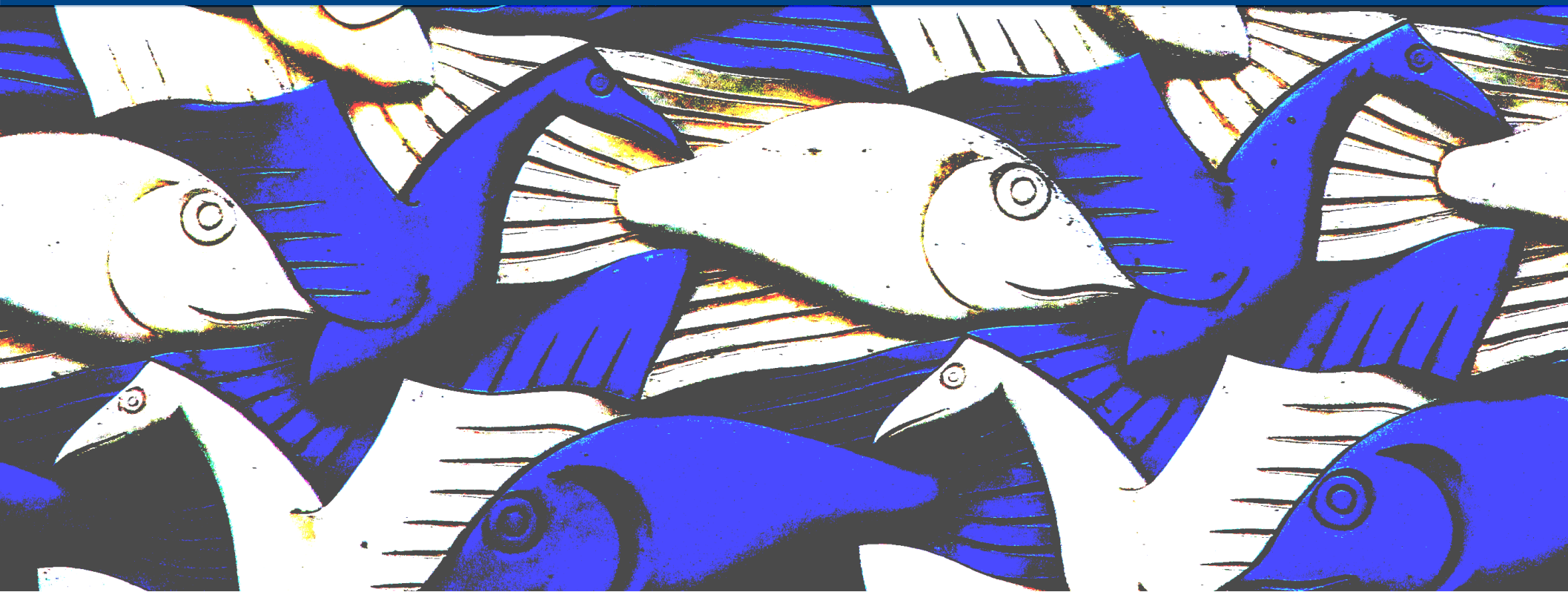


Analyse de réseaux sociaux



Raisonnement et science de la décision

Master M1 MIAGE – Ingénierie Métier

Université Toulouse 1 Capitole

Umberto Grandi

L'expérience de Milgram

En 1967, Stanley Milgram à Harvard fait une expérience pour détecter la “distance” entre deux individus moyens des États-Unis

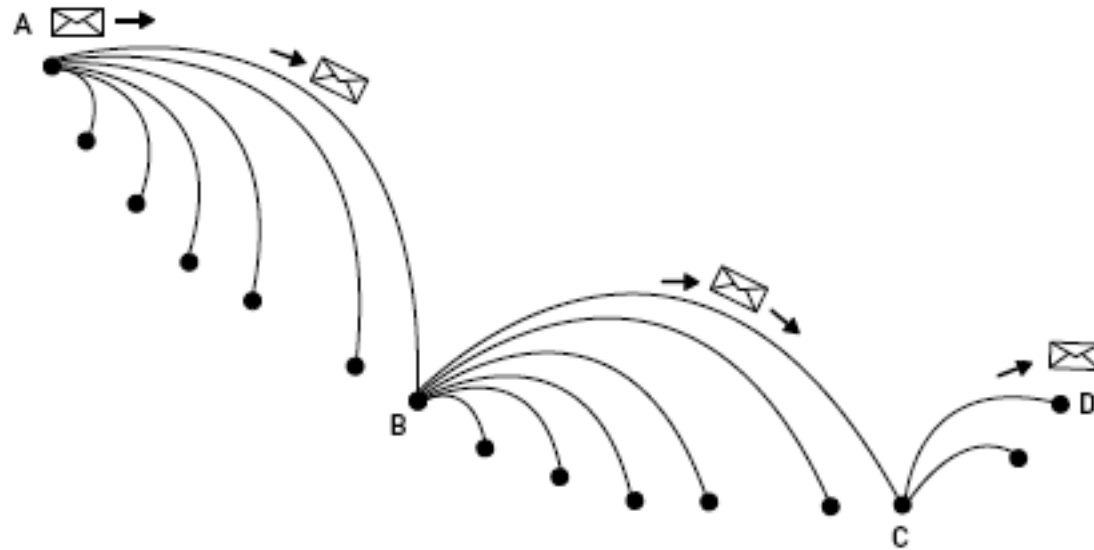


Une lettre était envoyée à des citoyens d'Omaha en Nebraska, avec le nom d'une personne de Boston. La lettre priait le citoyen d'Omaha de transférer le message directement au Bostonien si son adresse était connu, ou de la transférer à un ami ou proche qui, selon le receveur, pourrait connaître le Bostonien.

Combien de passages avant que la lettre arrive au Bostonien ? 5, 50, 100, 1000 ?

Six degrés de séparation

Entre 5,5 et 6.

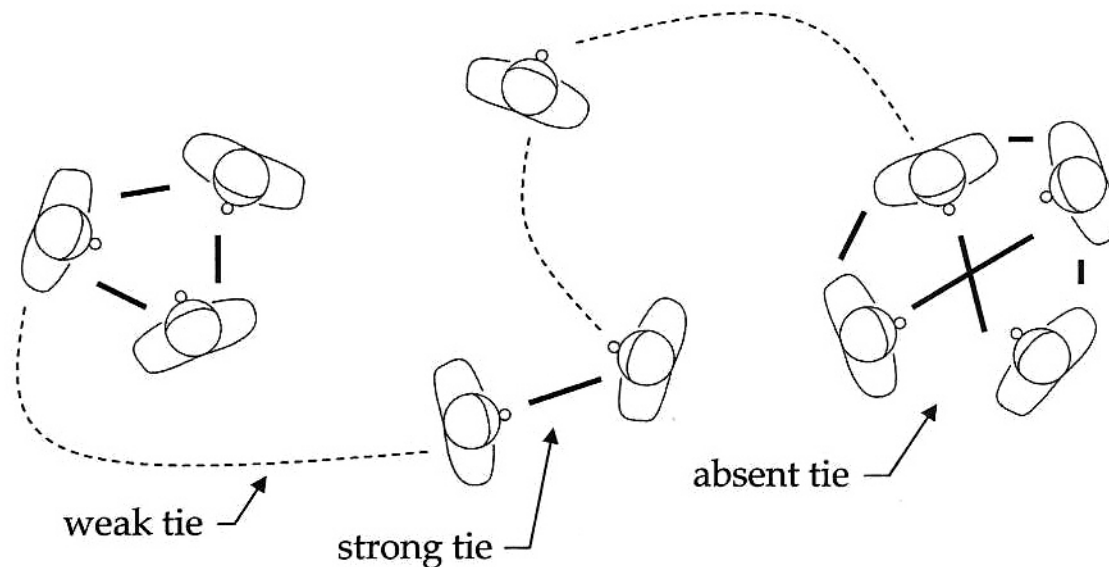


Sa conclusion est que les réseaux des connaissances ont une **longueur moyenne de connexion** qui est petite par rapport au nombre des nœuds :

- Les nœuds sont les personnes
- Les flèches ou arcs sont les liens d'amitié ou de connaissance
- Le réseau présente l'**effet de petit monde** (*small world networks*)

Une deuxième expérience

En 1973 Mark Granovetter recueille des données pour découvrir comment des personnes ont trouvé leur poste (*Linkedin* n'existait pas...mais quelle est la différence entre réseaux sociaux et média sociaux ?)

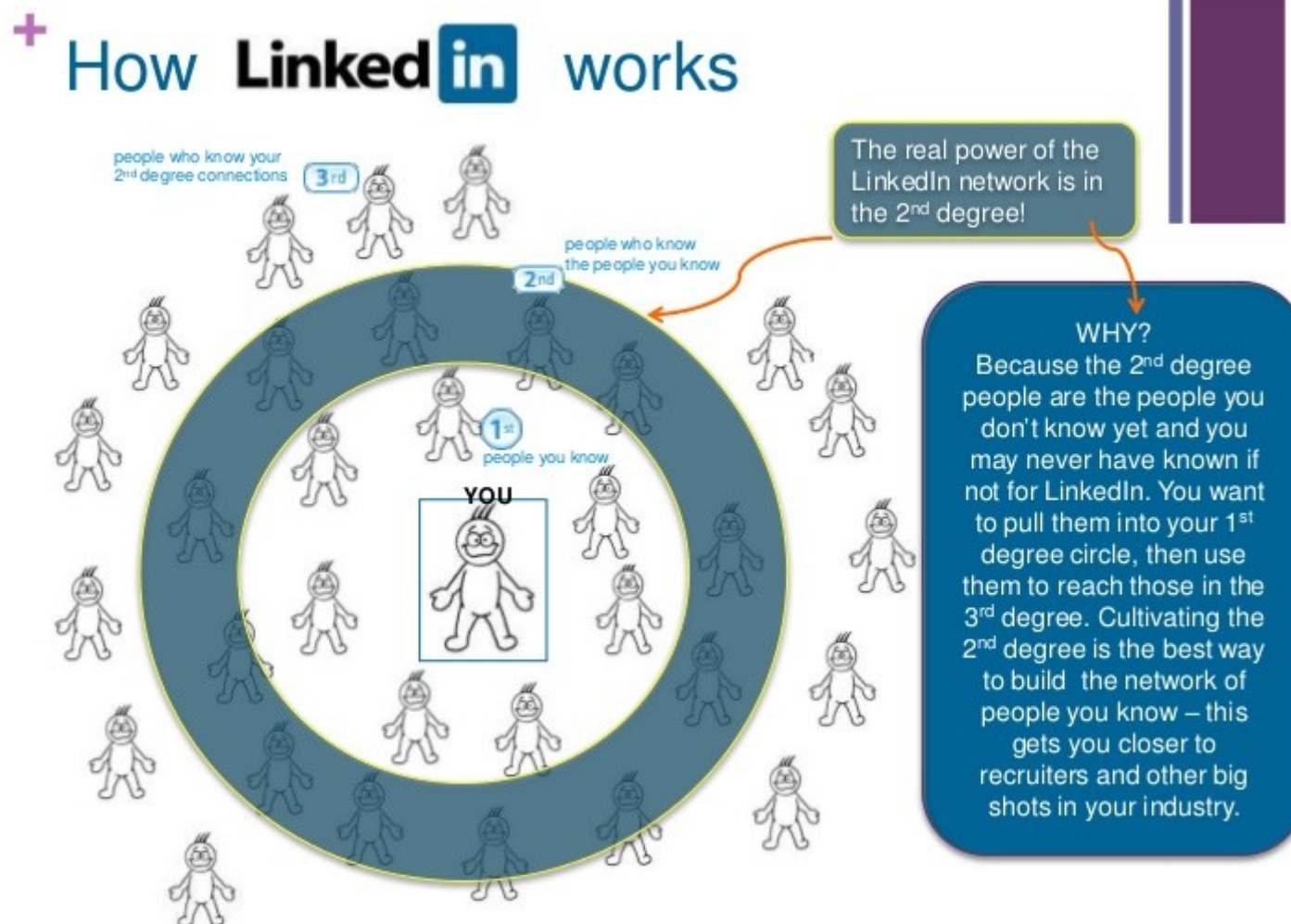


- La plupart des personnes ont utilisé leur réseau social pour trouver un poste
- Le poste n'a pas été trouvé par des amis mais par des **connaissances**

Cela montre l'importance des **liens faibles** dans les réseaux « petit monde » pour passer d'un cercle très connecté à un autre.

Pourquoi cela est important pour moi ?

Image de <http://www.slideshare.net/yaamary/using-linkedin-to-land-your-first-job-6770392>



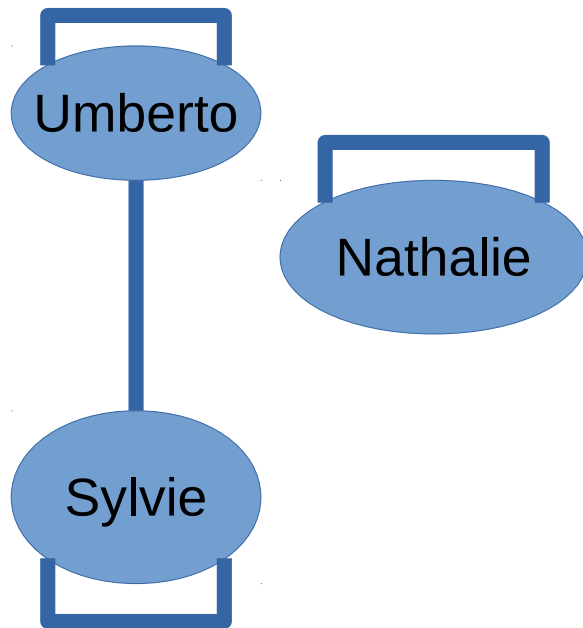
Avec seulement 6 recommandations vos CVs peuvent arriver sur le bureau de Mark Zuckerberg !

Ce cours vous donnera les instruments de base et le vocabulaire pour comprendre la **structure** d'un réseau (social). En particulier :

1. La représentation d'un graphe/réseau comme une **matrice d'adjacence** ;
2. Les notions de diamètre d'un graphe et de longueur moyenne de connexion, pour évaluer la **dimension** d'un réseau;
3. Deux **mesures de centralité** qui permettent d'identifier les nœuds le plus importants sur le réseau.

Premier définition : un réseau

Définition : un **réseau** ou **graphe** est donné par un ensemble de nœuds ou sommets et par une matrice d'adjacence g .



Nœuds : {Umberto, Sylvie, Nathalie}

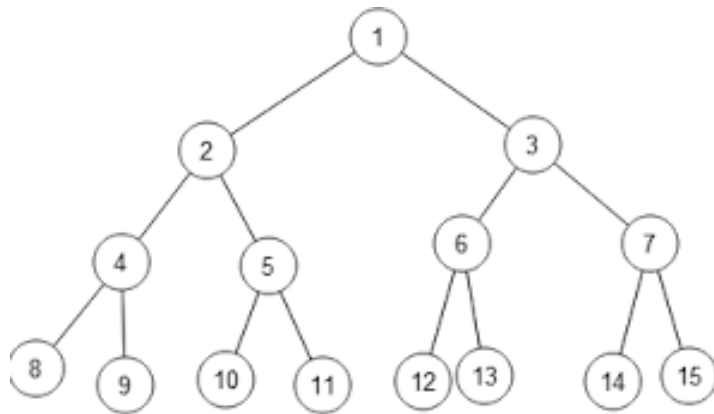
Matrice d'adjacence de la relation
« être dans le même bureau de » :

	U	S	N
U	1	1	0
S	1	1	0
N	0	0	1

Dans la matrice d'adjacence on écrivera 1 dans la position (x,y) si le nœud x est connecté à y . Par default on considérera des **graphes non-orientés** (arrêtes sans direction) : *quelle propriété a donc la matrice d'adjacence ?*
Quelle différence avec les réseau des pages WEB du dernier cours ?

Qu'est-ce que c'est ...

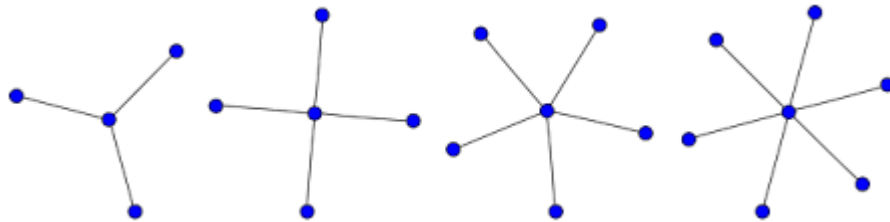
...un arbre ?



Un arbre est un graphe connecté sans cycles

Qu'est-ce que c'est...

...une foret ?

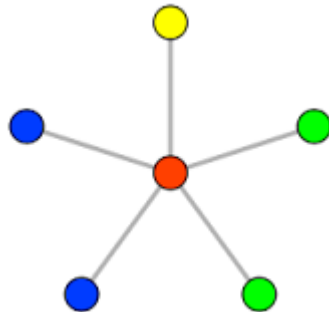


Une **composante connecté** est un ensemble de nœuds tel que chaque pair de nœuds est connecté par un **chemin**.

Une **foret** est un graphe tel que chaque composante connecté est un arbre.

Qu'est-ce que c'est...

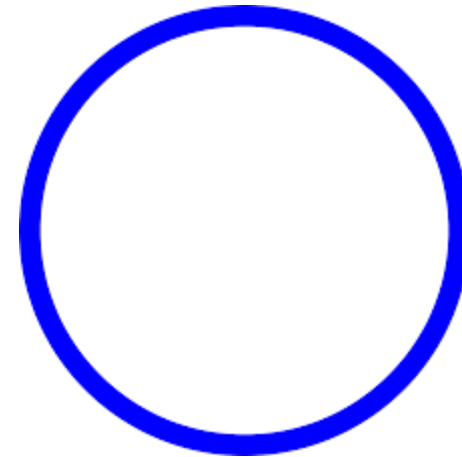
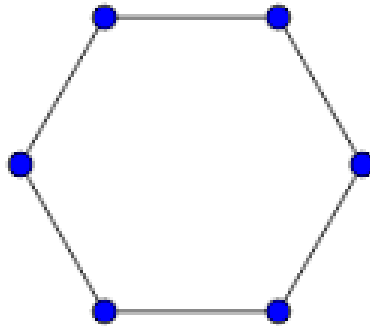
...une étoile ?



Une **étoile** est un réseau avec un noeud i tel que $g(i,j)=1$ pour tous noeds j (different de i), et tel que $g(j,j')=0$ autrement.

Qu'est-ce que c'est...




...un cercle ?



Un **cercle** (ou cycle simple) est un réseau avec un cycle et tel que chaque nœud est connecté à exactement deux autres nœuds.

Comprendre la structure d'un réseau

Soit (N,g) un réseau, et i un nœud :

- | | | |
|---|---|------------|
| 1) Quelle est la popularité de nœud i ? |  | Dégré |
| 2) Quelle est la dimension du réseau (N,g) ? |  | Diamètre |
| 3) Quelle est l' importance du nœud i ? |  | Centralité |

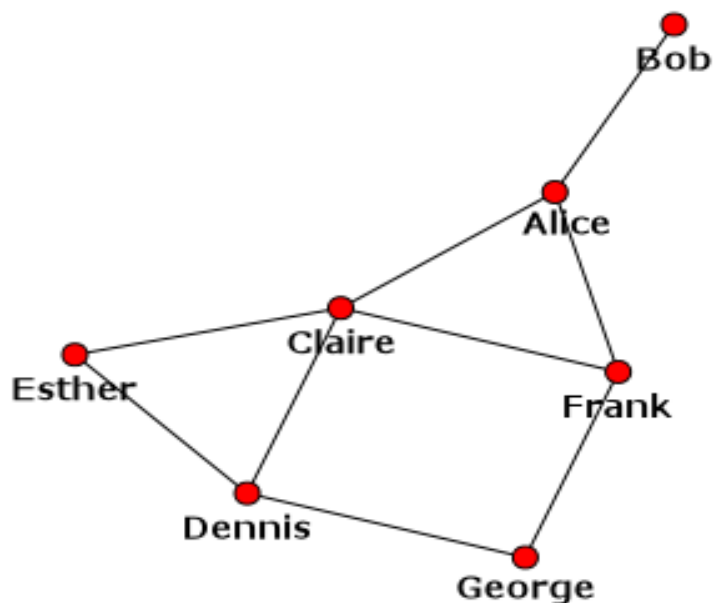
Allons donc voir les définitions formelles de ces trois notions...

Dégré et distribution de degré

Le **degré** d'un nœud est le nombre des nœuds qui lui sont connectés

Exercice : comment calculer le degré du nœud i sur la matrice d'adjacence ?

La **distribution de degré** $p(d)$ d'un réseau (N, g) est la fréquence (une liste de valeurs, ou une distribution de probabilité) des nœuds avec degré d .

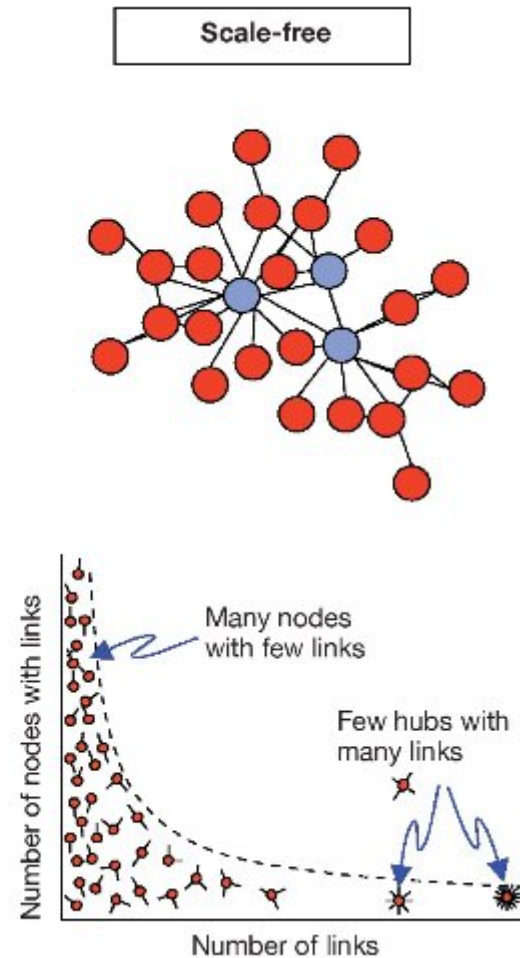
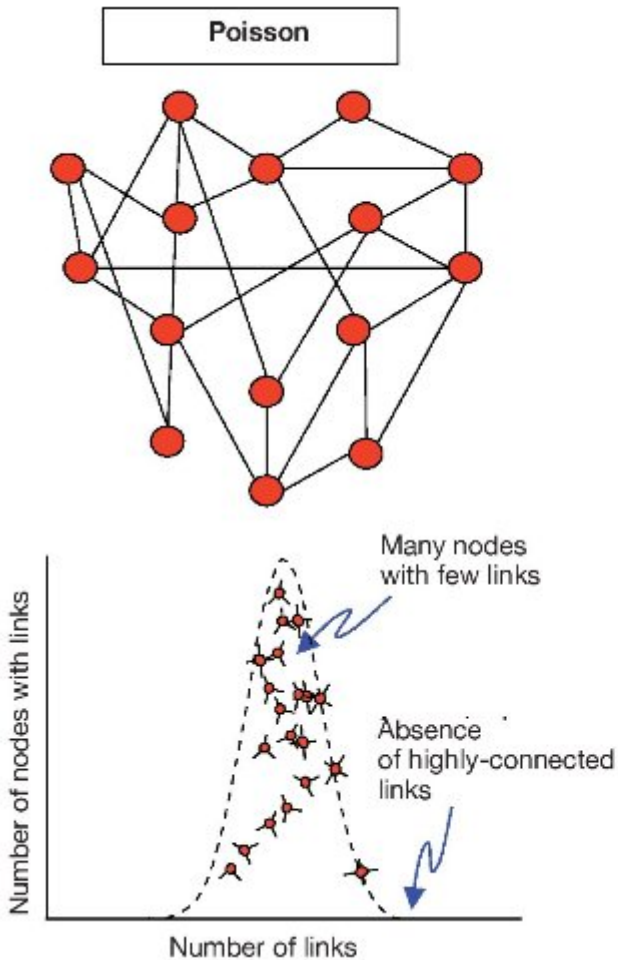


Distribution de degré

Degré	Nombre de nœuds
1	1
2	3
3	3
4	1

*Il y a un **erreur** dans la table, trouvez-le !*

Deux importantes distributions de degré (extra)



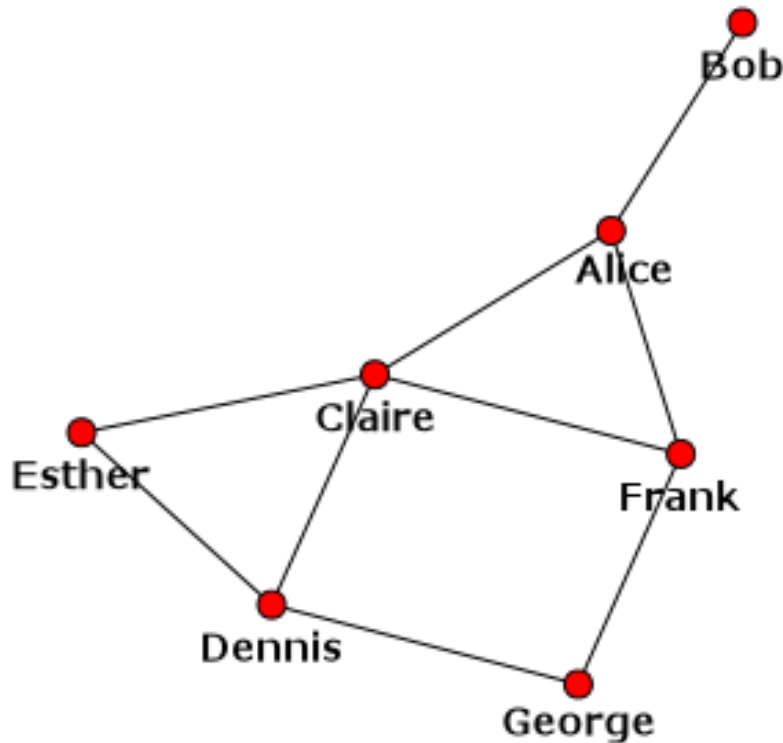
Exemple

Distribution de degré

Les réseaux générés de façon **purement aléatoire** ont une distribution de degré de Poisson, avec la plupart des nœuds avec peu de liens. Réseau dits **scale-free**, générés avec un algorithme dit « preferential attachment », ont une longueur moyenne de connexion basse : ils sont des réseaux petit monde !

La distance sur un réseau

La **distance entre deux nœuds** est *la longueur du plus court chemin* qui relie les deux nœuds. Si deux nœuds ne sont pas connectés leur distance est $+\infty$ (infinie).



Quelles est la distance entre Bob et George ?

Entre Esther et Frank ?

Du point de vue computationnel, est-il difficile de calculer la distance entre deux nœuds ?

Algorithme de distance de Dijkstra

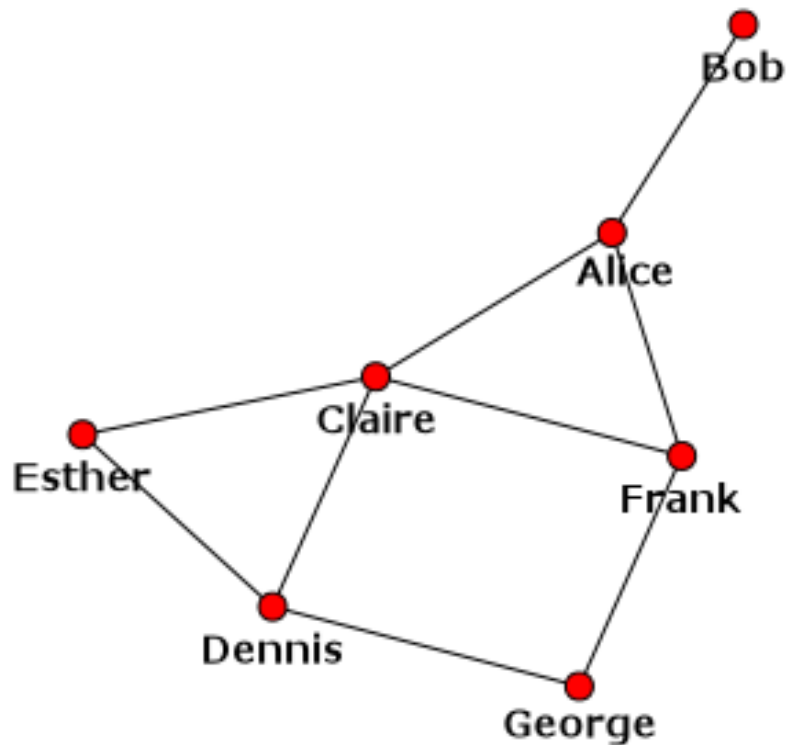
Le problème est similaire à un problème de routing : trouver la route plus courte pour un voyage en avion entre Toulouse et Minneapolis : pas de vols directs, mais d'abord Toulouse-Paris et après Paris-Minneapolis...

Algorithme de distance entre deux nœuds B et A (Dijkstra)

- 1) Initialiser $dist(A)=0$ et tous autres nœuds $dist(C)=infini$
- 2) Marquer A comme nœud *courant*, et tous autres comme *non-visité*
- 3) Considérer un voisin C du nœud *courant*. Comparer $dist(C)$ avec $dist(courant)+1$, et ré-initialiser $dist(C)$ au **minimum** de ces deux distances. Répéter pour tous voisins de *courant*.
- 4) Une fois tous voisins du nœud *courant* visités, marquer *courant* comme *visité*.
- 5) Si $courant=B$ c'est fini : la distance entre A et B est $dist(B)$
- 6) Autrement, sélectionner entre les nœuds *non-visité* celui avec $dist$ **minimale**, et le marquer comme *courant*. Retourner au pas 3)

Le diamètre d'un réseau

Le **diamètre** d'un réseau est la distance maximale entre deux nœuds.



Quel est le diamètre ?

D'un point de vue computationnel, est-il difficile de calculer le diamètre d'un graphe ?

Comment trouver le diamètre d'un arbre ? d'un cercle ?

Longueur moyenne de connexion

Rappelons-nous de l'expérience de Milgram des 6 degrés de séparation. La mesure utilisé n'était pas le **diamètre** (la **longueur maximale des distances**) mais la **longueur moyenne des distances** entre les nœuds.

Voici quelques observations issues du monde réel :

- Dans les réseaux de connaissances (Milgram, 1967) la longueur *médiane* de connexion est de 5,5.
- Dans le réseau de collaboration entre mathématiciens (expérience de Grossmann, 2002) la longueur moyenne de connexion est de 7.6 et le diamètre de 27.
- Facebook a une longueur moyenne de connexion de 4,74 (2012)
- Internet a une longueur moyenne de connexion de 3.1 (1999) – *Pourquoi Internet est un réseau ? Qui sont les nœuds ?*

Comment utiliser ces notions pour décider qui sont les nœuds les plus importants dans un réseau ?

L'origine du pouvoir des Medici à Florence

Ceci est le réseau de relations des familles nobles à Florence au 15ème siècle :
Pourquoi les Medici étaient si puissants ?

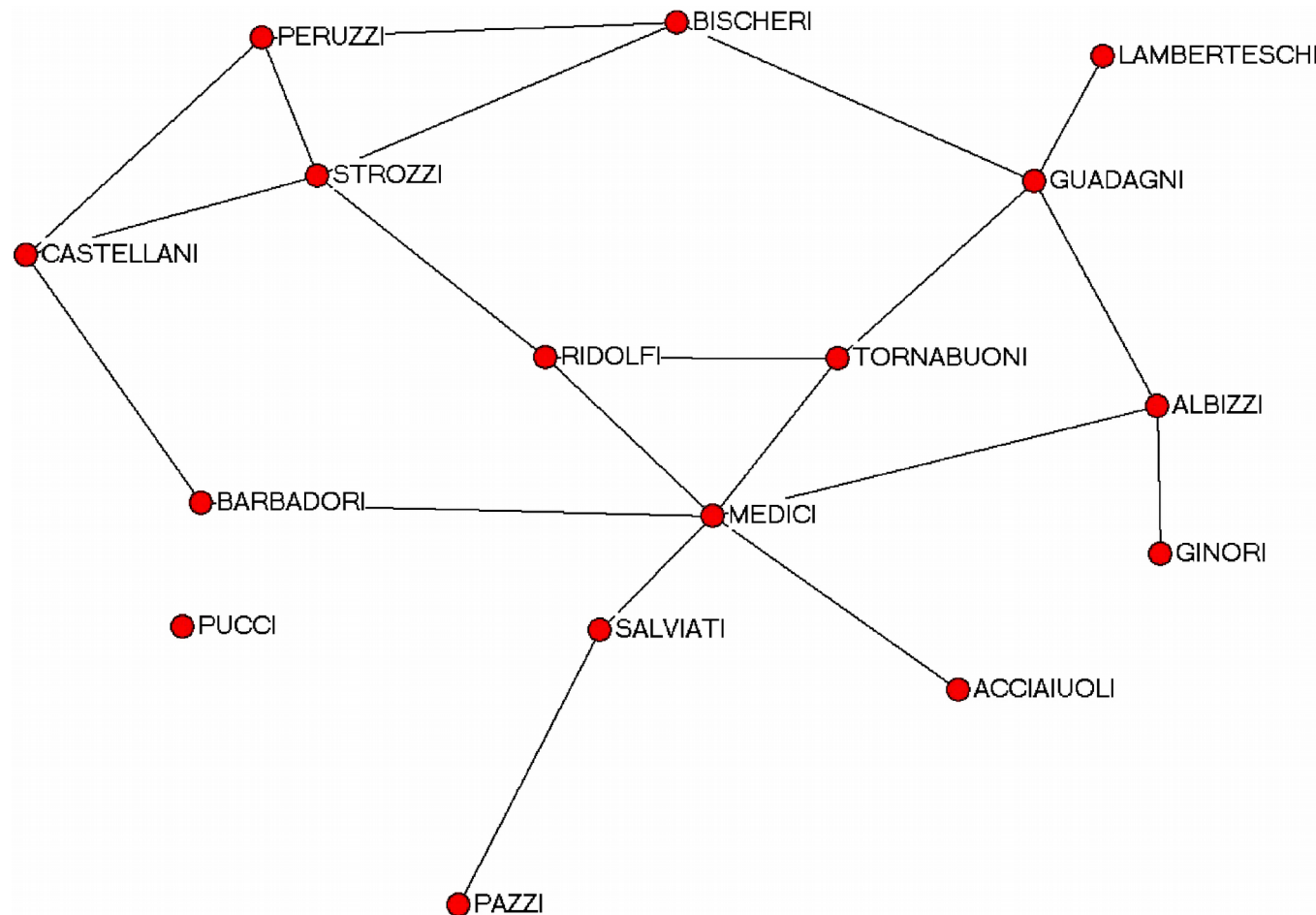
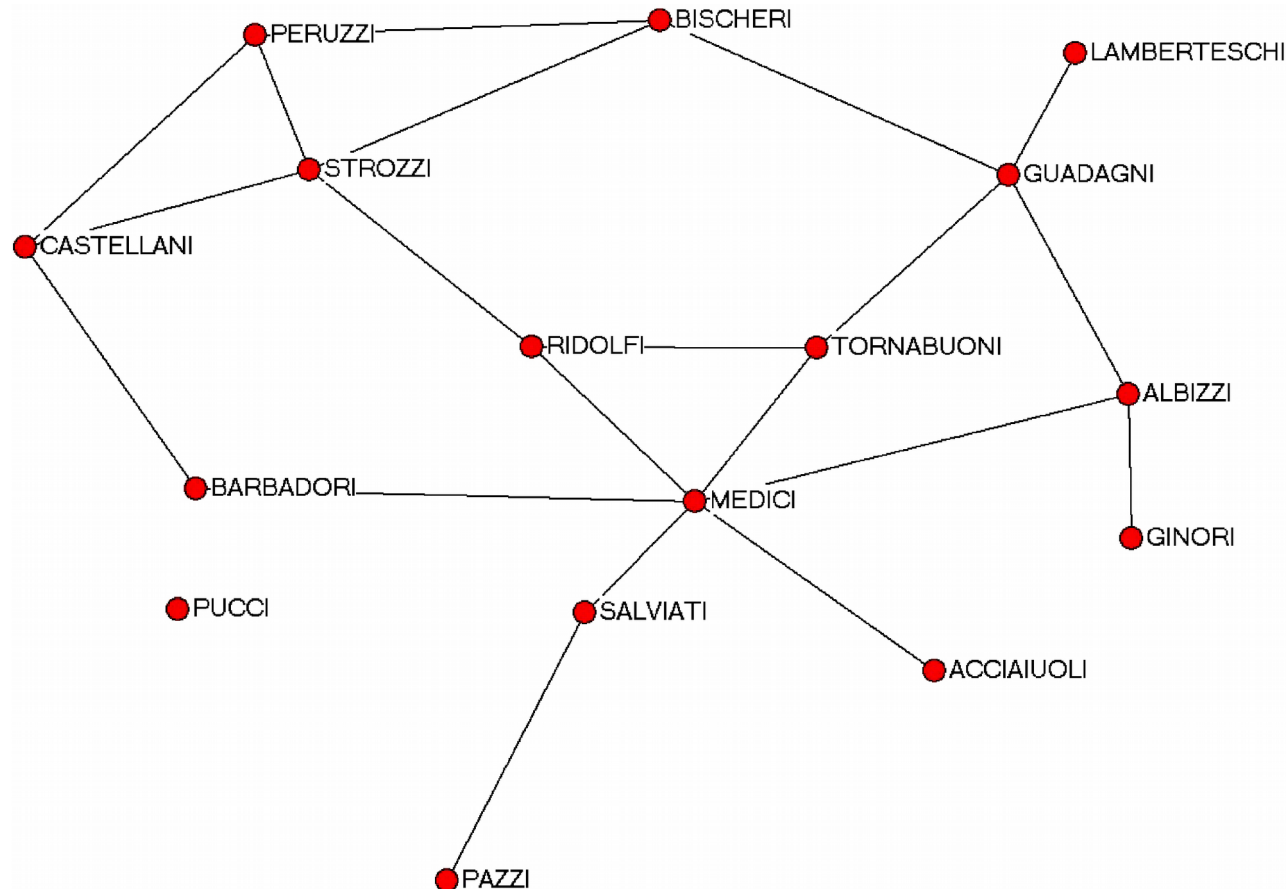


Figure de M. Jackson, Social and Economic Networks.

Structure de pouvoir : nombre de liens

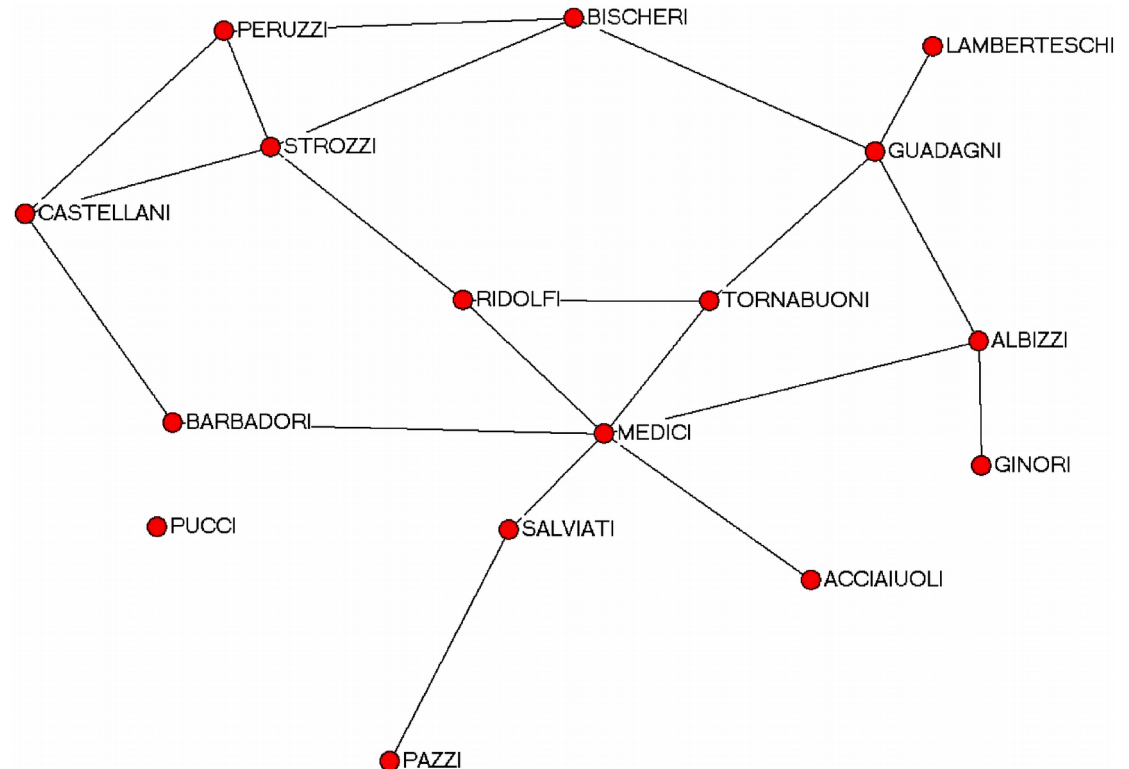
Sont ils **connectés** avec la plupart des autres familles?



Les Medici sont connectés avec 6 familles (ils ont un degré de 6), les Strozzi avec 4, les Guadagni aussi... **Le degré est une bonne mesure du pouvoir**, mais insuffisante pour expliquer le pouvoir des Medici !

Structure du pouvoir : conter les chemins

Si une famille B veut demander un faveur à une famille A à qui elle n'est pas directement connectée, il faut **passer par toutes les familles qui sont sur le chemin qui relie B à A...**



L. C. Freeman a proposé une **mesure de pouvoir** fondée sur cette notion : compter combien de paires de familles sont obligées passer par le Medici pour se connecter entre eux.

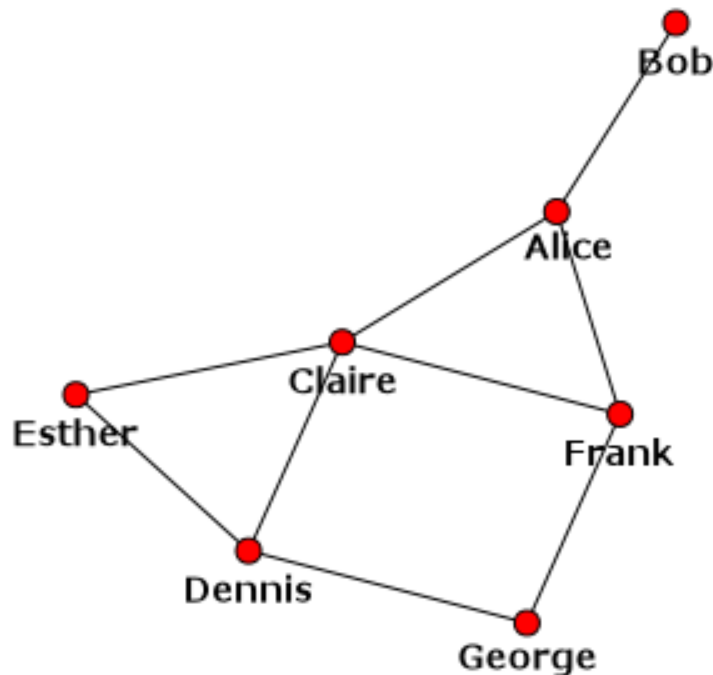
Avec cette mesure les Medici sont à 0,522 , les Strozzi à 0,103 et les Guadagni à 0,255. Ce qui explique mieux la structure de pouvoir existante !

Une première mesure de centralité

On cherche à définir et à trouver quels sont les nœuds les plus importants...plusieurs définitions sont possible pour mesurer la centralité/importance d'un nœud.

La **centralité de degré** d'un nœuds est mesuré par son degré.

Effectivement les degré plus connectés peuvent dans certains cas être les plus importants. Mais cela est une mesure trop locale !



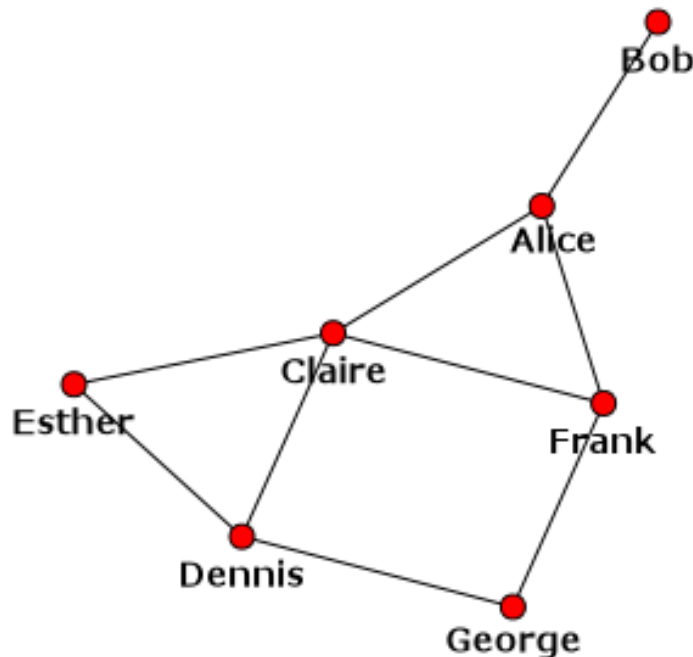
Qui est le nœud le plus important ?

D'un point de vue computationnel, est-il difficile de trouver le nœud le plus important ?

Une deuxième mesure de centralité

On a donc besoin d'une notion plus globale. Voici une mesure qui regarde tout le réseau pour décider de l'importance d'un de ses nœuds :

La **centralité de proximité** d'un nœud i est l'inverse de la distance moyenne de connexion entre i et tous autres nœuds.



*Quelle est la centralité de proximité de Claire?
Et de Frank ?*

*D'un point de vue computationnel, est-il difficile
de trouver le nœud le plus important ?*

La centralité de proximité en formule

La centralité du nœud i est calculée comme :

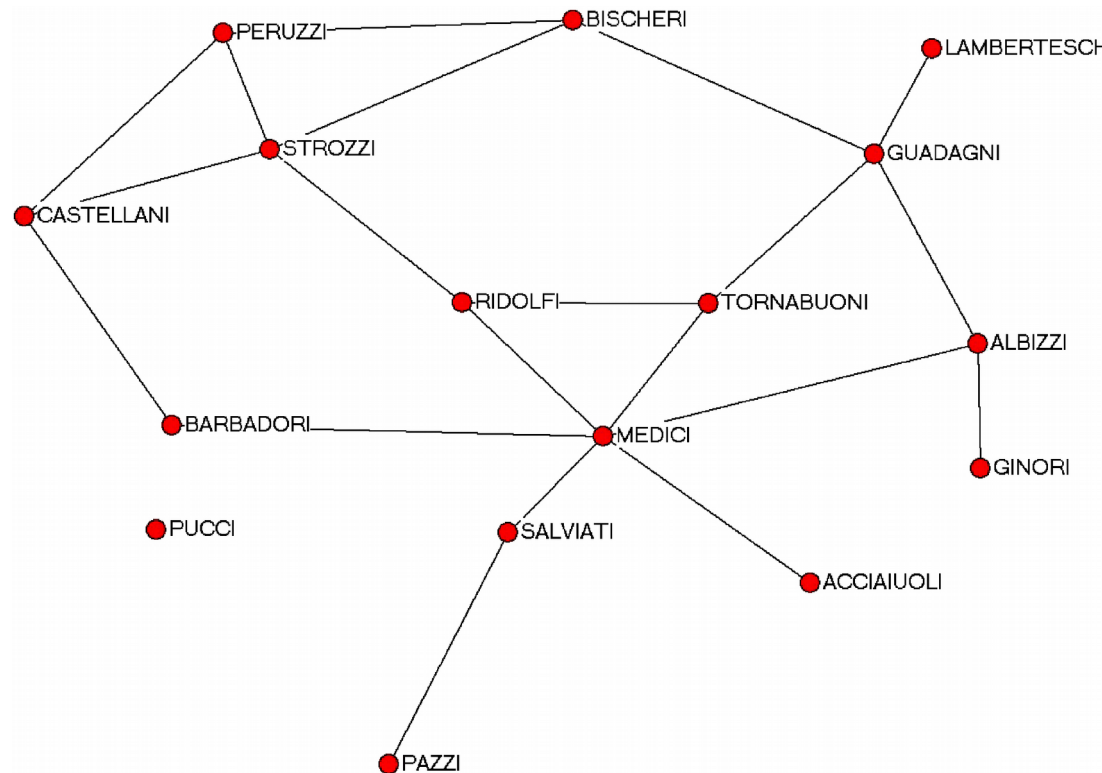
$$prox(i) = \frac{1}{\left(\frac{\sum_{j \in C(i)} dist(i,j)}{|C(i)|} \right)}$$

$$C(i) = \{j \mid i \text{ est connecté à } j\}$$

Observations :

- La **distance moyenne** de i à tous autres nœuds est calculé seulement avec les nœuds connectés
- Si i n'est pas connecté à personne cette formule ne marche pas (il faudra définir un cas particulier, par exemple donner un valeur de 0)
- *Exercice : écrire « i est connecté à j » formellement (en symboles)*

Centralité de proximité à Florence



Centralité de proximité des deux familles :

Medici	0,56
Strozzi	0,438

Conclusions

- Les réseaux sociaux sont omniprésents : pas seulement Facebook ou LinkedIn, mais les réseaux fournisseurs/clients, les réseaux familiales..
- L'analyse des réseaux ne s'applique pas seulement aux réseaux sociaux : Internet, les réseaux des ingrédients relié par des recettes, les gènes...
- *Exercice : dessinez une partie de votre réseau d'amis sur Facebook et analysez-le ! Pour les pros : téléchargez le réseau entier.*





- *Informatique décisionnelle (semestre 2)* : savoir **analyser un réseau d'employés** pour comprendre la structure de pouvoir, ou identifier des problèmes dans la chaîne de transfert d'information...
- *Stratégie (semestre 2)* : les médias sociaux ont souvent la forme d'un réseau, et savoir analyser quels sont les nœud plus importants est une tâche fondamentale en **marketing**.
- *Programmation structurée et objet (jusqu'en M2...)* : les algorithmes des graphes sont des excellents exercices pour **tester vos compétences en programmation !**

Ce cours est basé sur les premiers deux chapitres de :

- *Matthew O. Jackson. Social and Economic Networks. Princeton University Press, 2008.*

La vie de Stanley Milgram (mais pas l'expérience de six degrés de séparation) est décrite dans le film *Experimenter*, de Michel Almereyda, 2015.

Un bon livre de vulgarisation sur l'analyse des réseaux est :

- *Albert-Làszlo Barabasi. Linked : The New Science of Networks. Perseus, 2002.*