

Connectivity, Cycles, and Persistence Thresholds in Metapopulation Networks

JEAN-MARIE GOMÈS

21 novembre 2010



Article paru dans PLOS "Computational Biology" (2010), par Yael Artzy-Randrup et Lewi Stone.

Table des matières

Introduction	2
1 Modèle de base et notations	3
2 Programme de départ : une population isolée (n=1)	3
3 Modèle de n populations connectées	6
4 Influence de la topologie du réseau	7
4.1 Graphe régulier	7
4.2 Graphe aléatoire	7
4.3 Graphe hétérogène	7
5 Persistance et présence de cycles	8
5.1 L'intérêt de la dispersion est conditionné par l'existence de cycles . .	8
5.2 Les "arêtes isolées"	9
5.3 Les "arêtes critiques"	9
5.4 Etude de G cycle par cycle	9
6 Synthèse : cohérence de la population	10
7 Conclusion	11

Introduction

La conservation des écosystèmes, et donc la compréhension de leur dynamique, sont des thèmes centraux dans les défis du XXI^e siècle. L'écologie consiste en l'étude des interactions au sein d'un écosystème, et l'évolution des relations entre ses constituants au cours du temps ; en particulier, les observations de l'écologie peuvent être appliquées à la conservation des écosystèmes.

Dans l'article que nous allons étudier, les auteurs s'intéressent à la survie d'une métapopulation (réseau de populations en interaction), et en particulier à la dépendance de leur "seuil de persistance" en fonction de paramètres comme la connectivité, la présence et la nature de cycles dans le graphe représentant cette métapopulation. L'écologie moderne utilise des outils mathématiques puissants : ici, l'algèbre linéaire et la théorie des graphes, associés à la simulation informatique, permettent de dégager les éléments cruciaux pour le maintien de cette métapopulation.

De tels outils permettent parfois d'intervenir dans la dynamique des populations de façon éclairée et quantifiée, voire de mettre en évidence des vérités contre-intuitives (voir l'exemple de l'Ile de Santa Cruz, en Californie, où la cohabitation entre l'aigle royal et les renards, deux espèces protégées, était impossible).

On essaiera ici de comprendre en quoi l'utilisation de la théorie des graphes est pertinente en écologie ; pour cela, on se propose d'étudier le programme informatique employé par les auteurs. Toutefois, ce code n'étant pas publié, on propose ici une version extrapolée à partir des données de l'article. Ce programme manipule des notions telles que la connectivité dans un graphe, les graphes orientés et non-orientés, la présence et la nature de cycles...

Afin d'éclairer ces concepts et leur utilisation, on commencera par étudier un programme simple permettant de simuler la dynamique temporelle du nombre d'individus de la métapopulation. On complexifiera petit à petit ce code, en testant l'intérêt ou la validité de plusieurs hypothèses (hétérogénéité ou non du graphe, par exemple), en essayant de montrer à chaque étape quelle partie de la théorie des graphes est utilisée.

Mais avant de dévoiler le code, il apparaît indispensable de rappeler la structure de base du modèle et quelques notations.

1 Modèle de base et notations

La métapopulation est constituée de n populations, contenant des individus d'une même espèce regroupés en m classes d'âge. Le vecteur représentant la population i est noté $N^i(t)$ à la date t , et le j -ième coefficient de $N^i(t)$ est le nombre d'individus appartenant à la j -ième classe d'âge à la date t . La métapopulation est représentée par la matrice $m \cdot n$ notée $N(t)$, où la i -ième colonne représente la population i .

La matrice de Leslie associée au système est L , telle que : $N(t+1) = L.N(t)$. On note p_k , la probabilité de survie d'un individu d'âge $k-1$ jusqu'à l'âge k . f_k est le taux de fécondité de la classe k , alors que σ est la proportion de jeunes de la population qui est "auto-recrutée" avec succès après la phase de dispersion. On peut noter alors :

$$L = \begin{pmatrix} \sigma f_1 & \sigma f_2 & \sigma f_3 & \dots & \sigma f_m \\ p_1 & 0 & 0 & \dots & 0 \\ 0 & p_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & p_{m-1} & 0 \end{pmatrix}$$

Cette matrice peut être vue comme la combinaison linéaire d'une matrice "de survie" P et d'une matrice "de fécondité" F :

$$L = P + \sigma F$$

$$\text{avec } P = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ p_1 & 0 & 0 & \dots & 0 \\ 0 & p_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & p_{m-1} & 0 \end{pmatrix} \text{ et } F = \begin{pmatrix} f_1 & f_2 & f_3 & \dots & f_m \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 & 0 \end{pmatrix}$$

Les matrices P et F peuvent être distinctes selon chaque population, et seront alors notées P^i et F^i .

On note R , le "nombre de reproduction", représentant le nombre moyen de jeunes produits durant la vie d'un individu, défini par :

$$R = \sum_{k=1}^m f_k \prod_{l=0}^{k-1} p_l$$

À partir de ces outils de base, on peut étudier les conditions de la persistance de la métapopulation.

2 Programme de départ : une population isolée (n=1)

Le programme suivant, rédigé en langage Scilab, permet une simulation du modèle simple où il n'y a qu'une seule population ($n=1$).

```
clear
//Initialisation des paramètres
//Nombre de classes d'âge
m = 4;
//Nombre de populations
n = 1;
//durée de l'étude
duree = 12;
// Auto-recrutement :
```

```

sigma = 0.003;
//Probabilité de survie de chaque classe d'âge (très faible pour les jeunes,
élevé pour les anciens)
P = zeros(m,m);

for i = 2 : m
    P(i,i-1) = 10^(-m+i)*rand();
end

//Calcul du nombre de reproduction R (équation 2)
R = 0;
// Variable de calcul du produit des probabilités de survie
produit_des_p = 1;

for i = 1 : m
    R = R + F(1,i)*produit_des_p;
    if i < m then
        produit_des_p = produit_des_p*P(i+1,i);
    end
end

//Paramètre de persistance
khi = sigma*R;

//Structure initiale de la population : matrice n . m . duree
N = zeros(m,n,duree+1);
N(:, :,1) = int(100*rand(m,n));

//Et vogue la simulation...
//Initialisation du temps :
t = 0;

for t = 1 :duree
    //Application de l'équation (4) :
    N(:, :,t+1) = P*N(:, :,t);
    for i = 1 : n
        for j = 1 : n
            N(:,j,t+1) = N(:,j,t+1) + sigma*F*N(:,i,t);
        end
    end
end

//Représentation graphique de l'évolution de la population 1 au cours du
temps
//(et de ses différentes classes d'âge) : vecteur N1(i)

for i = 1 : m
    for t = 1 : duree+1
        pop1(t,i) = N(i,1,t);
    end
end

plot(pop1)

//affichage du paramètre de persistance
khi

```

Le programme renvoie une valeur : khi, qui intègre les informations sur R et σ , c'est-à-dire que l'efficacité "nette" de la reproduction est simplement le taux de reproduction par individu multiplié par la proportion de larves qui donneront réellement des adultes dans la population. La population se multiplie selon une loi géométrique de paramètre χ .

Lors du test de ce programme, en utilisant des valeurs réalistes pour les taux d'auto-recrutement, de fertilité et de survie, on obtient des valeurs de χ distribuées essentiellement entre 0 et 3. On observe que pour $\chi < 1$ (Exemple en figure 1a), la population ne survit pas à long terme : il s'agit d'une série géométrique de paramètre strictement compris entre 0 et 1, qui converge vers 0. Pour $\chi > 1$ (Exemple en figure 1b), la population subit une croissance exponentielle (au moins pour l'une de ses classes d'âge), correspondant à la divergence d'une série géométrique de paramètre strictement supérieur à 1.

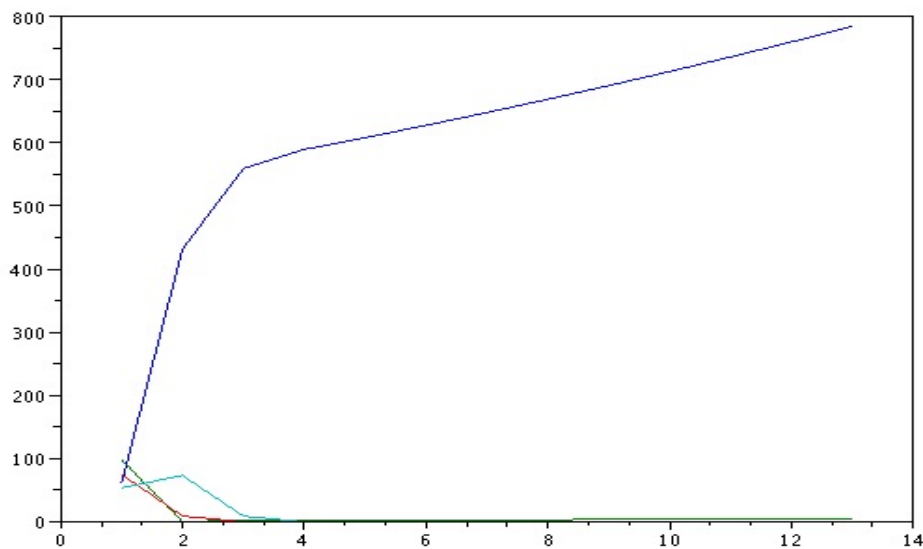
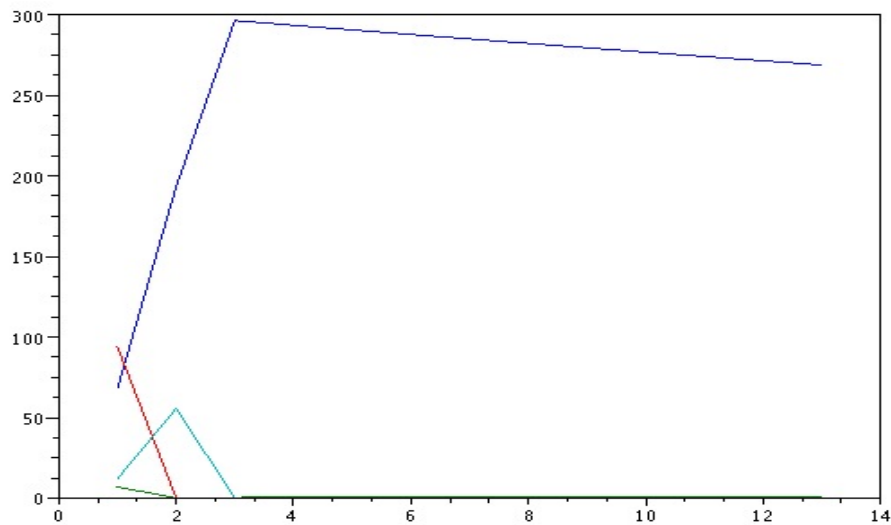


FIGURE 1 – Evolution temporelle de la structure de la population au voisinage de $\chi = 1$ ($\sigma = 0,03$; $a : \chi = 0,99$; $b : \chi = 1,03$). Les différentes courbes représentent l'évolution des différentes classes d'âge ; en particulier, la courbe bleue représente la première classe d'âge, largement dominante numériquement.

On appelle donc χ le **paramètre de persistance**, car sa valeur renseigne

directement sur la capacité de la population à se maintenir. On note : $\chi = \sigma \cdot R$. Existe-t-il un paramètre aussi informatif pour n populations interconnectées, et comment le calculer ?

3 Modèle de n populations connectées

Les auteurs poursuivent l'étude en créant une métapopulation, représentée par un graphe G dont les sommets sont les n populations, et les arêtes sont les flux de larves. Quelques hypothèses de travail sont proposées :

- Les matrices de fertilité et de survie sont identiques pour toutes les populations;
- La matrice de connectivité C est définie par $C_{i,j}$ = proportion de jeunes produits en j qui sont recrutés avec succès par la population i
 - Sans hypothèse supplémentaire, $\chi = \lambda_C \cdot R$, où λ_C est le rayon spectral de C (soit sa plus grande valeur propre)
 - Les auteurs font l'hypothèse que le nombre de jeunes qui immigrent dans une population j est en moyenne une proportion α de leur population d'origine j.

On peut alors considérer G comme un graphe non-orienté non pondéré; sa matrice d'adjacence A ($a_{ij} = 1$ s'il existe une voie de dispersion directe entre i et j, $a_{ii} = 0 \forall i$) est donc symétrique réelle. En particulier, la matrice de connectivité s'écrit sous la forme : $C = \sigma I + \alpha \lambda_A$, où I est la matrice identité, et λ_A le rayon spectral de A. Dans ces conditions, le paramètre de persistance s'écrit :

$$\chi = (\sigma + \alpha \lambda_A) \cdot R$$

On peut en tirer quelques conclusions immédiates et plutôt intuitives : si toutes les populations sont des "puits" (id est, ne peuvent pas se renouveler sans apport de larves extérieures : $\sigma R < 1$), c'est la structure du réseau qui détermine la survie de la métapopulation, via le rayon spectral A. Une dispersion importante et efficace se traduit par un paramètre α élevé, et aide au maintien des effectifs.

Le code peut être complété et modifié comme suit :

```
//Nombre de populations
n = 8;
//Proportion d'individus recrutés depuis chaque autre population :
alpha = 0,01/n;
//Structure de la métapopulation (matrice d'adjacence) : symétrique, coefficients inclus dans [0; 1]
A = zeros(n,n);
i = 0;
j = 0;

for i = 1 : n;
    for j = 1 : i-1
        A(i,j) = floor(2*rand());
        A(j,i) = A(i,j);
    end
end

//Calcul de la matrice de connectivité
C = sigma*eye(n) + alpha*A;

//Paramètre de persistance : max(spec(A)) est la plus grande valeur propre de A (réelle car A est symétrique réelle)
khi = sigma + alpha*max(spec(A));
```

//Et vogue la simulation...

$$(\dots) \quad N(:,j,t+1) = N(:,j,t) + C(i,j)*F*N(:,i,t); (\dots)$$

Les auteurs se proposent ensuite d'étudier plus finement l'influence de la structure du réseau sur la persistance de la métapopulation.

4 Influence de la topologie du réseau

Les auteurs se concentrent sur la répartition des liens, selon trois modalités principales : G peut être un graphe régulier, ou un graphe aléatoire homogène ou hétérogène.

4.1 Graphe régulier

Pour une topologie régulière, tous les sommets ont le même degré k (par exemple : $n = 8$, $k = 2$ pour un C_8 , voir Figure 2a). Le rayon spectral vaut alors k , et on a $\chi = (\sigma + \alpha k)R$. Les auteurs envisagent le cas où $\alpha = \sigma$: alors, l'auto-recrutement devient faible devant l'importance de l'immigration, et le degré des sommets contrôle principalement la persistance de la métapopulation.

4.2 Graphe aléatoire

Les auteurs utilisent un modèle de graphes aléatoires dû à Erdős et Rényi. Dans ce modèle, il y a une probabilité p pour que deux sommets donnés soient reliés par une arête, ce qu'illustre la figure 2b. On obtient aisément le degré moyen des sommets : $\bar{k} = n.p$. De plus, le rayon spectral de A est égal au degré moyen, d'où : $\chi = (\sigma + \alpha \bar{k})R$

4.3 Graphe hétérogène

On peut étendre ces résultats à tout graphe aléatoire, en quantifiant leur hétérogénéité. Pour ce faire, on peut utiliser la variable aléatoire k , dont les réalisations k_i correspondent au degré de chaque sommet. Le carré du coefficient de variation de k , noté CV^2 et défini par $CV^2 = Var(k_i)/\bar{k}^2$, est choisi pour représenter l'hétérogénéité du réseau. Par exemple, le réseau présenté en figure 2c est fortement hétérogène (un sommet très connecté, et la plupart des autres non).

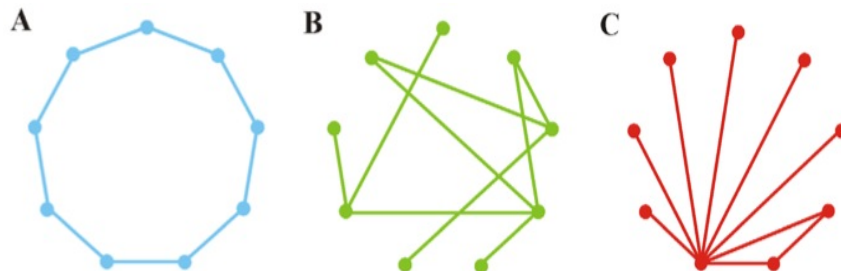


FIGURE 2 – Trois grands types de topologie du réseau : Régulier (A), Aléatoire (B), Hétérogène (C). (Figure 2 de l'article original)

Le choix de cette mesure donne l'expression du paramètre de persistance :

$$\chi = [\sigma + \alpha \bar{k}(1 + CV^2)]R$$

. Qualitativement, on peut interpréter cette formule de la façon suivante : la plus forte valeur propre de A dépend directement du degré du sommet le plus connecté,

d_{max} . On peut s'attendre à ce que d_{max} soit de l'ordre de la moyenne + la variance divisée par la moyenne, soit $\bar{k} + Var(k_i)/\bar{k}$.

On propose d'implémenter ces modifications de la façon suivante :

```
//Calcul du nombre moyen de liens par sommet "degre_moyen"
degre_moyen = sum(A)/n;
//Calcul de la variance des coefficients de A
VarA = 0;

fori = 1 : n;
    VarA = VarA + (sum(A(i, :))-degre_moyen)^ 2;
end

VarA = VarA/(n-1);
//Calcul de l'hétérogénéité (coefficient de variation au carré)
CV2 = VarA/(degre_moyen)^ 2;
//Paramètre de persistance
khi = (sigma + alpha*degre_moyen*(1+CV2))*R;
```

Plus G est hétérogène, plus la persistance est forte. Ce résultat, plutôt contre-intuitif, pourrait s'interpréter par la présence de quelques populations à fort nombre reproductif, et fort degré, dont les émigrants suffisent à maintenir toutes les populations qui leurs sont connectées.

Après cette étude générale, les auteurs s'intéressent à des propriétés plus fines de G : présence de cycles, liens critiques...

5 Persistance et présence de cycles

5.1 L'intérêt de la dispersion est conditionné par l'existence de cycles

On considère une métapopulation ne comprenant aucun cycle dans la dispersion. On souhaite savoir si cela a une conséquence directe sur le rayon spectral de la matrice d'adjacence, et donc sur le paramètre de persistance. Pour cela, on recherche les valeurs propres de A, ce qui se ramène à calculer les racines du polynôme caractéristique de la matrice $A' = (A - \lambda I)$, noté $P_A(\lambda)$. Les coefficients diagonaux de A' valent $-\lambda$ par définition, et les autres coefficients valent 0 ou 1. La formule de Leibniz donne :

$$P_A(\lambda) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^n a'_{(i, \sigma(i))}$$

Avec :

- σ est une permutation appartenant à S_n , l'ensemble des permutations de (1, 2, ... n)
- $\epsilon(\sigma)$ est la signature de chaque permutation ($\epsilon = +1$ si la permutation est paire, -1 si la permutation est impaire)
- $\sigma(i)$ représente l'indice qui est permuté avec i.

Pour toute permutation σ , avec $a'_{(i, \sigma(i))}$ valant 0, 1 ou λ , on a deux cas :

$$\prod_{i=1}^n a'_{(i, \sigma(i))} = 0$$

ou

$$\prod_{i=1}^n a'_{(i, \sigma(i))} = (-\lambda)^{n-m},$$

m étant le nombre d'indices pour lesquels $a'_{(i,\sigma(i))} = 1$, c'est-à-dire qu'il existe une arête entre les sommets i et $\sigma(i)$. Or, toute permutation peut être décomposée en un produit de cycles de permutations disjointes. Ainsi, dans le produit étudié, il existe une suite d'indices $(x_1, x_2, x_3, \dots, x_m)$, telle que $a_{x_k, x_{k+1}} = 1 \forall k$, et $a_{x_m, x_1} = 1$. S'il n'existe pas de cycle dans le graphe, alors

$$P_A(\lambda) = \prod_{i=1}^n a'_{(i,i)} = (-\lambda)^n$$

Les racines de $P_A(\lambda)$ sont toutes nulles : le rayon spectral de A est donc nul. Le paramètre de persistance redevient $\chi = \sigma R$, on se ramène au cas de n populations isolées : la dispersion est inutile s'il n'existe pas de cycles dans le réseau de populations.

5.2 Les "arêtes isolées"

Si l'on supprime toutes les arêtes de G qui ne font pas partie d'un cycle, les valeurs propres ne sont pas modifiées, et donc le paramètre de persistance reste identique, comme en figure 3 de l'article :

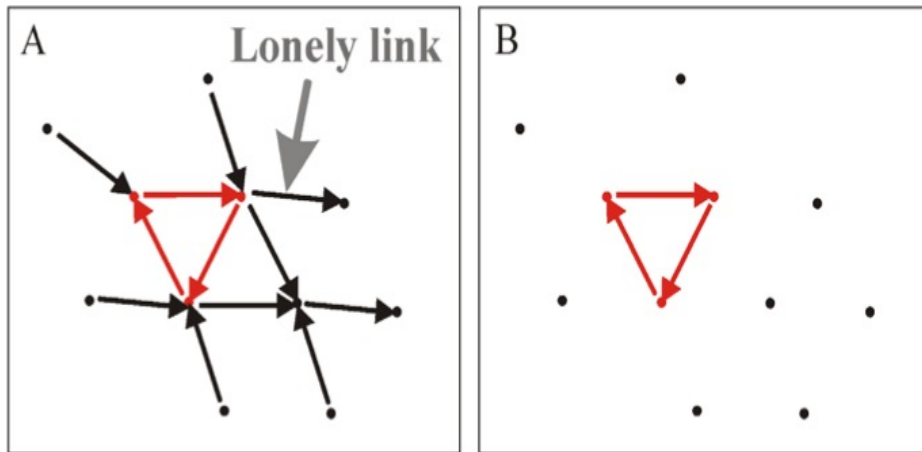


FIGURE 3 – Les deux matrices d'adjacence de ces graphes partagent le même polynôme caractéristique $P_A(\lambda) = \lambda^8 \cdot (\lambda^3 - 1)$. (Figure 3 de l'article original)

Ces voies de dispersion entre populations ne sont pas nécessaires à la survie de la métapopulation. Toutefois, elles peuvent être cruciales dans le maintien d'un nombre maximal de populations, et donc dans le maintien d'une forte diversité génétique, ce qui est très recherché en conservation des espèces.

5.3 Les "arêtes critiques"

Le pendant des arêtes isolées sont les arêtes critiques : si on peut supprimer un cycle de G en supprimant une seule arête, la préservation de celle-ci doit être prioritaire, comme en figure 4.

5.4 Etude de G cycle par cycle

Trois résultats majeurs se dégagent alors :

- L'importance du "composant cyclique dominant", c'est-à-dire du plus long cycle, est cruciale car il correspond au rayon spectral ; si ce cycle est brisé, le rayon spectral devient alors la valeur propre associée au deuxième plus long cycle. Ceci tend à diminuer χ , comme en figure 4.

- Les auteurs proposent alors d'étudier la persistance de la métapopulation en la séparant en composants cycliques disjoints, en particulier en éliminant les "arêtes isolées".
- Les auteurs remarquent que si le graphe n'est pas orienté, chaque paire de sommets connectés forme un "cycle de taille 2", ce qui rend la persistance plutôt robuste à la destruction d'arêtes. Si la dispersion de larves est "orientée", par exemple par la topographie ou par des courants marins, cela peut diminuer la persistance.

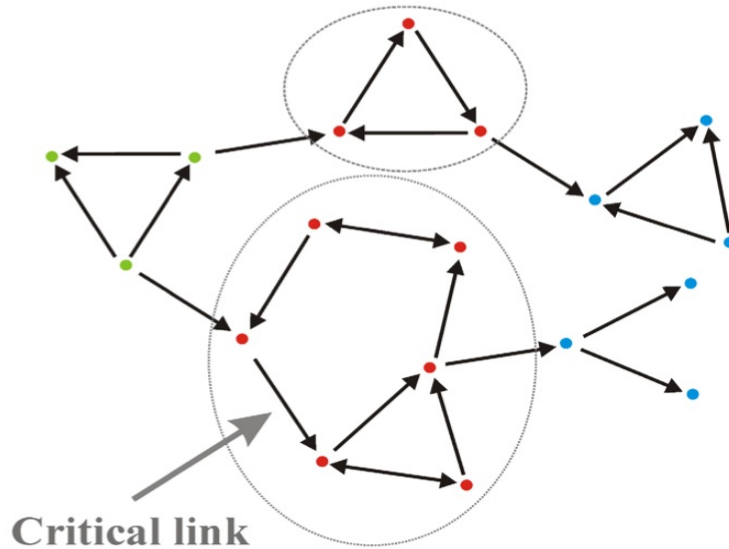


FIGURE 4 – L'arête indiquée par la flèche grise est critique ; la stabilité de la métapopulation passe par celle des composants cycliques.

6 Synthèse : cohérence de la population

Si une seule population est capable de se croître, alors toutes celles qui lui sont connectées, directement ou non, seront capables de survivre aussi. Plus le réseau de dispersion est fragmenté en sous-réseaux indépendants, moins cette propriété sera exploitée : les auteurs introduisent alors la notion de "cohérence" de la métapopulation. La cohérence recouvre la connexion à grande échelle des sous-réseaux. Elle permet par exemple d'amortir les effets des perturbations extérieures (notion de résilience en écologie), favorise la recolonisation des sites impactés... Une forte cohérence correspond donc à une mutualisation des risques sur l'ensemble de la métapopulation. Toutefois, si les évolutions temporelles des populations sont trop fortement corrélées, le risque d'extinction totale augmente.

Une valeur critique de p , la probabilité qu'il existe une arête entre i et j , est souvent retenue pour l'émergence d'un composant cyclique d'une taille de l'ordre de $n : p = 1/n$.

7 Conclusion

Dans cet article, les auteurs ont montré qu'il était possible de trouver un paramètre pertinent, permettant d'évaluer correctement la survie d'une métapopulation. La définition de ce paramètre varie selon les cas : le nombre de populations, le degré moyen des sommets de G , la topologie du graphe, son hétérogénéité...

Ce travail montre la puissance de la modélisation et des mathématiques, mais aussi ses limites :

- Si l'on veut appliquer ce modèle à une métapopulation réelle, il peut être délicat voire impossible d'évaluer certains paramètres de la modélisation ;
- Le modèle repose nécessairement sur des hypothèses simplificatrices (le taux de dispersion est identique pour toute paire de populations connectées), et néglige les interactions inter-spécifiques ainsi que la capacité du milieu ;
- La persistance, si elle est un premier objectif, est souvent conditionnée par la survie d'un maximum de populations, pour des raisons de diversité génétique ; c'est alors la "cohérence du réseau" qui intervient. Cette cohérence est plus difficile à quantifier, bien que la taille du composant cyclique maximal en soit un indicateur intéressant.

Cette étude met également en lumière le fait suivant : la science est capable d'empêcher l'effondrement de certains écosystèmes et la disparition de certaines espèces (bien qu'il soit trop tard pour beaucoup d'entre elles), problèmes très actuels et cruciaux pour la qualité de vie (voire la survie) de centaines de millions de personnes. Mais il est du devoir des scientifiques de rappeler à tous que la plupart de ces problèmes sont reliés directement aux activités humaines, et à une foi excessive en la technologie que la science elle-même a trop longtemps portée.