

Les réseaux sociaux dans la lutte contre le Covid-19

par Gianluca Manzo

Les prédictions de contamination du COVID-19 sur lesquelles reposent les décisions de santé publique négligent les apports des sciences sociales, notamment l'analyse des réseaux sociaux qui pourrait contribuer à affiner les modèles de diffusion permettant d'armer la lutte contre la pandémie.

Contexte et objectifs

Dans un article récemment publié dans *La vie des idées* (<https://laviedesidees.fr/Savoir-et-prevoir.html>), Pascal Marichalar étudie les articles de synthèse sur le COVID-19 parus dans la revue *Science* entre début janvier et début mars ainsi que les publications de l'Organisation Mondiale de la Santé. Sa conclusion est sans appel : les autorités sanitaires et politiques françaises ne pouvaient pas ne pas « savoir ». La critique est tout aussi percutante : elles auraient dû « prévoir ».

Cet essai porte sur ce deuxième aspect¹. Il reprend la chronologie des faits à partir du moment où les décideurs politiques ont intensifié leurs activités de prévision. Le 11 mars, jour où 2281 cas de COVID-19 sont désormais enregistrés en France par l'Agence Nationale de la Santé, à la demande du président de la République, le ministre de la santé et des solidarités

¹ Plusieurs collègues m'ont aidé à améliorer ce texte en prenant le temps de le commenter par écrit. Je voudrais remercier en particulier Mohamed Cherkaoui, Pierre Demeulenaere, Michel Dubois, Olivier Galland, Gérald Bronner, Jean-François Mignot, Lucas Sage et Dominique Vidal. La rédaction de *La Vie des Idées* m'a également utilement poussé à clarifier et développer certains points. Je reste bien entendu le seul responsable du propos et des erreurs éventuelles qui y persisteraient.

forme un conseil scientifique avec l'objectif d'« éclairer la décision publique dans la gestion de la situation sanitaire liée au Coronavirus » (https://solidarites-sante.gouv.fr/IMG/pdf/200311-conseil_scientifique.pdf). Des restrictions des libertés individuelles d'une ampleur inédite se mettent alors progressivement en place pour freiner l'épidémie. J'étudierai ici un ingrédient spécifique de ce changement de régime : les modèles mathématiques utilisés pour prévoir l'évolution de l'épidémie.

Dès la publication le 17 mars d'un long article dans les colonnes du *New York Times* intitulé *Behind the Virus Report That Jarred the U.S. and the U.K. to Action*, on commence à voir la proximité temporelle qui existe entre le moment où les prévisions des modèles mathématiques sont présentées aux décideurs politiques et l'annonce publique de ces mesures, avec parfois, comme dans les cas américain et anglais, des revirements de stratégie spectaculaires. Depuis, les journalistes scientifiques continuent de recenser les modèles mathématiques que les experts présentent aux gouvernements en Europe et aux États-Unis. La revue *Science* (doi:10.1126/science.abb8814) le fait le 25 mars, suivie par *Nature* (doi: 10.1038/d41586-020-01003-6) le 3 Avril. La presse française ne fait pas exception. *Le Monde* consacre son dossier Science & Médecine du 1er Avril à la modélisation des épidémies. L'effort pédagogique de ces articles est remarquable. Une limite majeure des modèles présentés est cependant systématiquement sous-commentée : leur incapacité à traiter la structure des interactions sociales qu'ils supposent pourtant être les canaux de la transmission du virus. Cette simplification n'est pas anodine : elle conditionne le type de recommandations que nous pouvons formuler pour enrayer l'épidémie. C'est la thèse principale de ce court essai.

Pour la défendre, je commencerai par rappeler la séquence temporelle des recommandations formulées par le conseil scientifique et les choix du gouvernement ayant conduit au confinement généralisé. Je décrirai ensuite la nature des modèles mathématiques sur lesquels se sont fondées les recommandations en question. J'expliquerai dans un troisième temps comment ces modèles représentent les interactions sociales et les conséquences de ces choix de modélisation sur les interventions que nous (ne) pouvons (pas) dessiner. Je proposerai enfin les chemins que la puissance publique devrait emprunter pour donner aux chercheurs la possibilité de construire des modèles de diffusion des épidémies qui puissent prendre au sérieux la structure des interactions sociales².

Les avis du conseil scientifique et la prise de décision

Dès son *premier avis* du 12 mars, le conseil scientifique se réfère à des modèles mathématiques simulant l'évolution de l'épidémie pour montrer ce qui se passerait pour les hôpitaux si une réduction drastique des interactions sociales ne se réalisait rapidement (voir

² Tous les avis du conseil scientifique sont librement téléchargeables ici : <https://solidarites-sante.gouv.fr/actualites/presse/dossiers-de-presse/article/covid-19-conseil-scientifique-covid-19>.

notamment p. 2). Parmi les mesures conseillées pour atteindre cet objectif, figurent l'« isolement à domicile des personnes symptomatiques et des contacts intra-domiciliaires », le « confinement des personnes de plus de 70 ans, et des personnes médicalement fragiles », la limitation de « l'utilisation des transports en commun », la « fermeture des crèches, des écoles, des collèges, des lycées et des universités », la « mise en place du télétravail » et la « réduction des activités sociales et de loisirs non essentiels » (*ibidem*, p. 5). Le soir du jeudi 12 mars, le président de la république annonce la fermeture des établissements scolaires et exhorte la population à limiter le plus possible ses sorties et ses déplacements.

Le 14 mars le conseil scientifique formule un *deuxième avis* où « les modélisations épidémiologiques connues » sont à nouveau invoquées comme un élément à prendre en compte (voir notamment p. 1). Le conseil suggère de durcir la « restriction de la vie sociale ». Le soir du samedi 14 mars le premier ministre annonce la fermeture de lieux de loisirs, les commerces non essentiels à la vie du pays ainsi que l'interruption des offices religieux.

Le conseil scientifique rend un *troisième avis* le 16 mars. Il constate que « le comportement d'une partie des citoyens français ne s'est pas suffisamment modifié ». Des recommandations plus sévères apparaissent, notamment « la mise en œuvre d'un confinement généralisé strict sur le modèle de l'Italie, mis en place au niveau national sur l'ensemble de la population française, avec éventuellement des mesures plus strictes dans les zones les plus touchées » (point 1, p. 2). Les modèles mathématiques apparaissent à nouveau au cœur du raisonnement : « Une réduction d'au moins 60 % des contacts est nécessaire pour endiguer la dynamique de l'épidémie, comme le montrent certaines études de modélisation » (*ibidem*). Le soir du lundi 16 mars, le président de la république annonce l'installation du confinement généralisé et l'introduction de sanctions pour celles et ceux qui ne respecteraient pas cette mesure.

Le conseil scientifique formule un *quatrième avis* le 23 mars où le prolongement du confinement est considéré « indispensable », et devant durer « vraisemblablement six semaines au moins à compter de sa mise en place » (p. 2) ; d'autre part, le conseil « estime de manière consensuelle nécessaire un renforcement du confinement ». Cet avis ne mentionne pas les modèles mathématiques mais il est utile de rappeler que les simulations sur lesquelles le conseil s'est basées (à savoir celles de l'équipe de Neil Ferguson, voir section suivante) ont été produites sous l'hypothèse que le confinement reste en place pendant « *5 months or longer* » (voir doi:10.25561/77482 [p. 4]). Le soir du 23 mars le premier ministre annonce des restrictions supplémentaires aux déplacements des personnes ; le 27 mars le même premier ministre annonce le prolongement du confinement généralisé jusqu'au 15 avril au moins. Nous savons à présent (9 Avril) que cette mesure sera prolongée au-delà de cette date.

Ainsi, au vu de la séquence temporelle des événements, bien que le conseil scientifique précise que les résultats des modèles mathématiques « doivent être pris comme un élément parmi d'autres pour informer le décideur » (avis du 12 mars, p. 2), l'hypothèse que ces modèles aient fonctionné comme une boussole pour orienter les choix politiques paraît raisonnable. Cela

peut se justifier. Les rares évaluations *ex-post* auxquelles ont été soumis les modèles mathématiques que je vais décrire suggèrent que, tant qu'on se limite à une ou deux semaines, les prédictions macroscopiques du nombre de contaminations produites par cet outil sont relativement fiables (Funk et al. 2019). Le conseil scientifique a par ailleurs fait une utilisation prudente de ces prédictions en privilégiant notamment « des hypothèses intermédiaires faisaient déjà apparaître avec un degré de plausibilité élevé, un écart important entre le nombre de cas sévères nécessitant une réanimation avec les capacités hospitalières françaises, même augmentées par des mesures appropriées » (avis du 12 mars, p. 3). Ce qui est en cause n'est donc pas l'utilité prédictive de ces modèles qui fournissent une indication macroscopique grossière mais précieuse de la dynamique des contaminations.

La vraie question est en réalité ailleurs. Il s'agit de comprendre si, par-delà la prédiction, ces modèles mathématiques sont aussi utiles pour nous indiquer comment éteindre la dynamique de l'épidémie. Poser cette question est d'autant plus important que ces mêmes modèles fournissent désormais également le cadre général de la réflexion sur les rythmes du déconfinement, en France (Di Domenico et al. 2020 ; dans la presse, voir aussi *Le Monde* du 8 Avril, p. 2) comme à l'étranger (Kissler et al. 2020). Voilà donc le vrai enjeu. Ces modèles peuvent-ils nous aider à concevoir des interventions ciblant des (petits groupes d') individus infectés et des chemins particuliers reliant plusieurs individus infectés ? Ou bien ces modèles nous limitent à des recommandations générales touchant de manière indiscriminée des pans entiers de la population, voire celle-ci dans son ensemble comme dans le cas extrême du confinement ? Pour répondre à cette question, un détour par le contenu exact des modèles ayant guidé les décisions politiques est utile.

Les simulations de Neil Ferguson et les modèles à compartiments en épidémiologie

Les prédictions concernant le nombre de contaminations et de décès, avec ou sans des interventions drastiques de distanciation sociale, que le conseil scientifique a présentées au premier ministre et au ministre de la santé le 12 mars sont issues de simulations réalisées par l'épidémiologiste Neil Ferguson, directeur du *MRC Centre for Global Infectious Disease Analysis* de l'*Imperial College London*. À ce jour, cette équipe a rédigé quatorze rapports. D'après ce qui est écrit dans l'avis du conseil scientifique du 12 mars (p. 2), le « modèle COVID19 particulier » auquel se réfère le texte correspond vraisemblablement au modèle finalement publié par Ferguson et son équipe le 16 mars dans le rapport numéro n° 9³.

Dans ce rapport (voir en particulier p. 4-5), le modèle est présenté comme une « simulation basée sur les individus » et il est explicitement dit que cette simulation adapte au

³ Tous les rapports sont librement consultables ici : <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/>.

COVID-19 les valeurs des paramètres d'un modèle précédemment utilisé pour modéliser la diffusion de la grippe en Grande-Bretagne et aux États-Unis, modèle lui-même initialement appliqué à la transmission de la grippe H5N1 en Asie du Sud (ces travaux ayant été publiés dans la revue *Nature* respectivement le 27 juillet 2006 et le 8 septembre 2005). La structure du modèle remonte donc à il y a au moins quinze ans.

Le conseil scientifique a jugé que « ce modèle reste une référence pour la planification des pandémies. »

Le modèle de Ferguson représente donc une population de plusieurs millions d'individus dont la distribution de l'âge et la taille de la famille reproduisent les distributions observées au Royaume-Uni et aux États-Unis. Chaque individu occupe une position dans l'espace et il est assigné à une école et/ou à un lieu de travail, la taille et la distribution de ces deux lieux étant finement calibrées à travers des données administratives et géo-référencées pour reproduire la distribution géographique réelle des écoles et des lieux de travail. Au sein de cet environnement, chaque individu évolue au fil du temps à travers trois états possibles : sain, infecté et guéri. Le cœur du modèle concerne le mécanisme de transmission du virus d'un individu « infecté » à un individu « sain ».

Trois éléments entrent dans la représentation de ce mécanisme : 1/ le nombre de jours pendant lesquels un individu « infecté » peut réellement contaminer un individu « sain » (c'est la période dite d'« incubation ») ; 2/ la probabilité que, au sein de chaque interaction avec un individu « sain », l'individu « infecté » transmette effectivement le virus (les « infectés » étant à cet égard différenciés en « symptomatiques » et « asymptomatiques », les premiers étant supposés avoir une probabilité de contamination plus élevée que les seconds) ; 3/ le nombre de contacts sociaux de chaque individu. La simulation du rythme de l'épidémie au niveau macroscopique est enfin déterminée, entre autres, par les hypothèses sur le moment à partir duquel un individu « infecté » peut transmettre le virus (selon qu'il est « symptomatique » et « asymptomatique »), sur la proportion de cas qui requièrent une hospitalisation, sur la durée de celle-ci et sur la probabilité de décès en fonction de l'âge.

Le modèle que Ferguson a simulé pour évaluer l'impact des mesures de distanciation sociale sur la diffusion du COVID-19 en Grande-Bretagne et aux États-Unis est donc une forme raffinée d'une classe de modèles très répandus chez les biologistes et les épidémiologistes pour étudier les épidémies : les « modèles à compartiments » (pour une introduction, voir, par exemple, Keeling et Rohani 2008). Ces modèles étudient la transformation au fil du temps de la proportion d'individus appartenant à un nombre limité (défini au départ) de groupes, par exemple les individus « sains » (dits *Susceptible* dans la littérature anglo-saxonne), « infectés » (*Infected*) ou « guéri » (*Recovered*). Puisque ces trois groupes sont très souvent utilisés, on parle aussi de modèle SIR (étiquette que j'utiliserai dans la suite pour abrégé). La particularité du modèle de Ferguson est de proposer une version d'un modèle SIR où le mécanisme de transmissions du virus est modélisé au niveau des individus. Le plus souvent les modèles à compartiments sont en revanche étudiés à l'échelle des groupes en décrivant l'évolution de leur

taille au fil du temps à travers la technique des équations différentielles. Ceci peut expliquer d'ailleurs une partie du succès de ces modèles (j'évoquerai en conclusion un deuxième élément concernant les données empiriques). Ils sont à la fois simples et explorables à travers un formalisme mathématique puissant et bien maîtrisé. C'est un mécanisme connu. Des modèles peu réalistes peuvent néanmoins durablement dominer un domaine de recherche, car ils sont supportés par un appareillage mathématique commode (que l'on pense, en économie, aux modèles d'équilibre générale, fondés sur la simplification de l'agent représentatif, critiquée par Alan Kirman ou encore, en finance, aux modèles de prédictions de la valeur des actifs boursiers, ignorant eux les événements rares, au cœur de travaux de Christian Walter).

Ferguson lui-même, dans le rapport numéro 12 publié le 23 mars, change d'ailleurs de niveau d'agrégation pour étendre la simulation de l'épidémie du COVID-19 à 202 pays. Dans ce cas, n'ayant pas accès, pour tous les pays, à la même richesse d'informations empiriques, il formule un modèle SIR plus classique qui confirme, pour la Grande-Bretagne et les États-Unis, l'ordre de grandeur des prédictions (du nombre total de contaminés et de décès en l'absence de toute intervention) produites deux semaines plus tôt à travers les simulations centrées sur les individus. Que les variantes microscopiques des modèles SIR demandent des données plus riches est un fait connu : quand ces données existent, leur valeur ajoutée est de permettre des prédictions plus fines à des échelles plus petites telles qu'une région, une ville ou un groupe d'âge donné, par exemple (pour ce type de comparaison, voir Ajelli et al. 2010).

Les modèles SIR sont en général présentés comme un outil pour décrire une quantité épidémiologique fondamentale, à savoir le taux de reproduction, noté R_0 : il représente le nombre moyen d'individus qu'un individu infecté peut contaminer (tant qu'il est contagieux). C'est un paramètre fondamental car il est censé nous renseigner sur le *seuil critique*, à savoir le point de basculement au-delà duquel la diffusion du virus ne peut plus être arrêtée (j'y reviendrai). Or, ce taux est en réalité le résultat de trois composantes que la simulation individu-centrée de Ferguson permet de distinguer clairement au niveau microscopique, à savoir le temps pendant lequel un individu peut en infecter un autre, la probabilité qu'il l'infecte effectivement et le nombre de contacts sociaux d'un individu infecté. À l'évidence, ces derniers jouent un rôle central dans la dynamique du modèle : l'action des deux premiers paramètres se déploie en effet encore et encore à chaque interaction. Mais comment les modèles SIR représentent-ils ces interactions ?

Les interactions sociales dans les modèles à compartiments

Dans la micro-simulation de Ferguson publiée le 16 mars, chaque individu est supposé avoir des contacts au sein de sa famille, à l'école (si cela s'applique), sur le lieu de travail (si cela s'applique) ainsi que dans tout autre espace social, ce dernier type de contacts étant établis au hasard (*randomly*, écrivent les auteurs) bien qu'ils soient supposés se raréfier en fonction de la

distance physique entre les individus. De plus, pour les individus ayant été assignés à une école, il est supposé que leurs contacts sont deux fois plus nombreux à l'école que dans d'autres lieux ; dans la population dans son ensemble, la distribution des contacts est déterminée de telle sorte qu'un tiers des interactions ait lieu en famille, un tiers à l'école et sur le lieu travail et un tiers partout ailleurs dans l'espace social. Dans les modèles SIR agrégés développés pour étendre les simulations à 202 pays, Ferguson retient en revanche une seule caractéristique fondamentale : la structuration du nombre de contacts en fonction de l'âge des individus (voir le rapport n° 12 du 23 mars, figure 2d-f, p. 5). Or, bien que les mesures de distanciation sociale soient habituellement présentées comme visant génériquement la réduction du R_0 , ces interventions touchent en réalité une composante spécifique du R_0 , à savoir le nombre de contacts des acteurs. C'est d'ailleurs bien dans les termes de la proportion de contacts que telle ou telle intervention est censée détruire que les deux rapports de Ferguson et de son équipe s'expriment (voir, par exemple, le tableau 2 du rapport 9 du 16 mars ou les pages 6-7 du rapport 12 du 23 mars). Soit dit en passant : cela est surprenant. Après tout, on pourrait également simuler l'impact de l'adoption des gestes barrières en faisant varier systématiquement une autre composante R_0 , à savoir la probabilité de transmission au sein d'une interaction donnée. Il n'en est rien. Le nombre de contacts d'un individu est le seul paramètre qui est modifié de façon exogène. Comme l'admet Ferguson lui-même à plusieurs reprises le comportement des acteurs n'est pas explicitement représenté. Les modèles SIR et le formalisme des équations différentielles s'y prêtent mal (je reviendrai sur ce point dans la dernière section).

Quelles hypothèses sommes-nous donc en train de faire quand nous concevons une intervention sous la forme d'une réduction, par exemple, de 75 % de toutes les interactions en dehors du noyau familial (l'une des interventions étudiées par l'équipe de Ferguson correspondant au confinement généralisée) ?

Premièrement, les modélisateurs à la Ferguson font comme si, au sein de ce bassin d'interactions, tous les acteurs se valaient. C'est une simplification qui va à l'encontre d'une régularité formidable des réseaux sociaux, à savoir la distribution fortement hétérogène du degré des acteurs. Ceci veut dire que certains individus ont des dizaines de contacts alors que la plupart n'en a que très peu. Ces individus constituent ce que Barabasi (2014 : ch. 5) appelle *hubs* ou *connectors*. Il est vrai que, dans certains modèles SIR, comme nous venons de le voir, le nombre de contacts est supposé varier en fonction de certaines caractéristiques de l'individu, principalement l'âge : c'est une avancée notable. Le problème est que, au sein même d'un groupe d'acteurs donné, leur degré (c'est-à-dire, leur nombre de liens) peut être très hétérogène. Pour se limiter à un seul exemple bien documenté, à âge et genre égaux, quelques individus ont des dizaines, voire des centaines de contacts sexuels (sur une période donnée) alors que la plupart n'en ont que quelques-uns (Liljeros et al. 2001).

Une deuxième hypothèse sous-tend la manière dont les interactions sont traitées dans les modèles SIR et ses variantes raffinées, individu-centrées, comme celles utilisées par Ferguson et son équipe : la supposition que, étant donné le nombre de contacts d'un acteur, ses rencontres, au sein d'un sous-groupe donné (comme une tranche d'âge) et/ou au sein d'un

espace social donné (comme l'école) se fassent au hasard. Là encore c'est une simplification qui va à l'encontre d'une régularité formidable des réseaux sociaux : chaque individu n'a pas la même probabilité de rencontrer n'importe quel autre individu au cours de ses activités. En d'autres termes, les interactions sociales sont structurées et contraintes : elles ne suivent pas des chemins aléatoires.

Cela vaut tout d'abord pour les contacts de chaque acteur. Pour se limiter à l'un des faits les plus étudiés, si un acteur est en contact avec deux autres, la probabilité que ces deux individus soient eux-mêmes en contact est bien plus élevée que si les rencontres se faisaient au hasard (c'est la tendance vers ce qu'on appelle en jargon technique « fermeture triadique »). Cette régularité est très étroitement liée à une autre propriété lourde des contacts sociaux d'un acteur donné : la pression vers la construction de relations avec des individus qui nous ressemblent (ce qu'on appelle en jargon technique l'« homophilie »). Que l'homophilie, et la pression vers la fermeture triadique qui en découle, puisse accélérer la création de foyers locaux de contamination est un résultat connu (voir, par exemple, Jackson et Lopez-Pintado 2013). C'est que, si A et B ont un contact infecté C en commun, ils peuvent se contaminer mutuellement avec une plus haute probabilité qu'en l'absence de ce contact commun C. Cette propriété pourrait être exploitée pour identifier des chemins de diffusion locaux plus probables que d'autres. Il n'en est rien. Les modèles SIR ne s'intéressent pas à la structure locale des réseaux sociaux. Mais la combinaison de cette myriade d'amas de contacts autour d'un acteur donné (ce qu'on appelle en jargon un réseau « égocentré ») produit des structures à une plus grande échelle qui suivent aussi des régularités connues. En particulier, les travaux de modélisation des réseaux dit « petit-monde » ont montré que ces multiples « îlots » de liens denses autour de chaque acteurs sont reliés par un petit nombre de liens-ponts qui augmente fortement la probabilité que deux individus, n'ayant aucune chance d'interagir si les rencontres se faisaient au hasard, peuvent en réalité établir un contact : ce sont précisément ces liens-ponts qui peuvent altérer profondément l'évolution d'un processus de contamination, notamment en accélérant sa vitesse et son taux de pénétration (Watts et Strogatz 1998). Ces liens pourraient être exploités pour identifier des chemins faisant communiquer des foyers infectieux éloignés, et donc freiner la diffusion. Il n'est en rien. Les modèles SIR ne s'intéressent pas non plus à la structure globale des réseaux sociaux.

Or il importe de comprendre qu'il existe un lien étroit entre le fait que certains acteurs ont de très nombreux liens (alors que d'autres n'en ont que très peu), la pression vers la « fermeture triadique » des réseaux égocentrés et la structure du réseau résultante à grande échelle. Ce sont en effet précisément les *hubs*, pour reprendre le terme de Barabasi, qui, à travers leurs multiples liens, contribuent à la fois à créer une très grande densité de liens au sein des cercles d'interaction locaux (les réseaux égocentrés) et à mettre en relation ces foyers en principe séparés les uns des autres. Ce mécanisme donne aux réseaux sociaux (et pas seulement) leur forme typique en arbre ou, si l'on préfère, en archipel (Barabasi et Bonabeau 2003). C'est la raison profonde pour laquelle ces réseaux sont très robustes face à des perturbations qui toucheraient un très grand nombre de nœuds (individus) avec un faible degré (nombre de liens)

alors qu'ils s'exposent à toute sorte de cascades dès lors que quelques *hubs* seulement sont touchés (Barabasi 2014 : ch. 9).

Ainsi, s'il y avait une seule conséquence à tirer de la nature structurée des interactions sociales pour intervenir sur la diffusion d'un virus, elle serait la suivante : essayer d'identifier (et de tester/isoler/traiter, selon les cas), de façon systématique, les individus ayant la plus grande probabilité d'établir les plus grand nombre d'interactions sociales. Les *hubs* sont cruciaux à double titre : s'ils ne sont pas encore infectés, ils sont davantage exposés à la contamination car ils sont en contact avec un plus grand nombre d'individus ; dès lors les *hubs* sont en revanche infectés, ils peuvent infecter très rapidement un très grand nombre d'autres individus. C'est pourquoi les *hubs* doivent être la cible prioritaire de l'intervention. Couper leurs liens entraînent des effets systémiques dont l'ampleur n'est pas comparable à l'isolement d'un grand nombre d'individus ayant très peu de connections (voir Barabasi 2014 : ch. 10). C'est en effet la neutralisation des *hubs* qui conduit à la fragmentation du réseau, en contribuant ainsi à l'élimination du virus par étouffement de foyers infectieux locaux qui ne communiquent plus. Les « petits » nœuds n'ont qu'une incidence négligeable sur ce processus de fragmentation. C'est un fait théorique et empirique connus par ceux qui étudient et modélisent la diffusion des maladies sexuellement transmissibles dans des petits groupes : les acteurs (relativement peu nombreux) ayant un très grand nombre de relations en parallèle (ce qu'on appelle « *concurrency* » dans cette littérature) ont un impact décisif sur la diffusion du virus et doivent être ciblés en priorité (voir, par exemple, Morris et Kretzschmar 1997 ; Rocha et al. 2011 ; Moody and Benton 2016). Raisonner en termes de contacts moyens et d'interactions aléatoires, comme le font les modèles SIR les plus couramment utilisés, conduit à sous-estimer l'importance de ces individus et de ces chemins particuliers qui, en dépit de leur nombre relativement limité par rapport à l'ensemble des individus et des chemins existants, sont des vecteurs formidables de la diffusion du virus.

Bien que minoritaires pour l'instant, plusieurs études théoriques existent sur l'introduction de la topologie des interactions sociales dans les modèles SIR (pour une synthèse de la littérature, voir Duan et al. 2014 : 809-815). Elles montrent que, même dans ces modèles, prendre en compte la structure des interactions sociales a des implications radicales. Pour se limiter au résultat le plus spectaculaire, revenons au taux de reproduction R_0 . Ce taux correspond, rappelons-le, au nombre moyen d'individus qu'un individu infecté peut contaminer tant qu'il est contagieux. Le principe selon lequel ce taux doit descendre sous la barre de 1 (le *seuil critique* auquel je faisais référence plus haut) pour que l'épidémie du COVID-19 s'éteigne nous est répété quasi quotidiennement par les médias ; il fonde tous les modèles et toutes les analyses contrefactuelles de Ferguson et de son équipe que je viens de commenter. On passe en revanche plus souvent sous silence les conditions sous lesquelles ce principe a été démontré mathématiquement, à savoir que a/ tous les acteurs ont virtuellement le même nombre de contacts et une probabilité similaire de transmettre l'infection et b/ que ces acteurs interagissent de façon aléatoire dans l'espace social (c'est l'hypothèse que l'on qualifie de « *homogenous* » ou, parfois « *uniform* » *random mix*).

Or, les (rares) modèles SIR qui font tomber cette hypothèse et étudient l'évolution d'un virus au sein d'une population d'acteurs reliés par une structure des interactions en arbre ou en archipel, pour reprendre l'image adoptée plus haut, ont découvert qu'une valeur de R_0 en dessous de 1 n'assure pas l'extinction de l'épidémie (pour une synthèse sur ce point, voir Barrat et al., 2008 : ch. 9). Pourquoi ? La raison tient une fois encore à l'action potentielle des acteurs ayant un très grand nombre de connections, les *hubs*. Un nombre moyen d'individus infectés par un individu contaminé inférieur à 1 peut en effet tout aussi bien résulter de la présence, d'une part, d'une myriade d'individus ayant très peu de contacts (individus qui dépriment donc le nombre moyens de contaminations) et, d'autre part, la présence de quelques individus au nombre de contacts très élevé : selon la cartographie exacte des liens, il suffit que l'un d'entre eux soit contaminé pour que la dynamique s'enclenche ou pas (quelle que soit donc la valeur du R_0). C'est pourquoi la seule distribution du nombre de contacts n'est pas suffisante pour prédire le résultat systémique : la structure globale du réseau doit aussi être connue (Barabasi 2014 : ch. 10).

En un mot, les modèles qui jouent le rôle principal dans l'aide à la décision dans la crise actuelle font l'hypothèse que les interactions entre les individus sont au cœur du processus de diffusion du virus, mais ils font le choix technique de ne pas modéliser explicitement la structure de ces interactions, ni au niveau local (les réseaux égocentrés) ni à l'échelle du système social (la topologie du réseau global). Pour utiliser une comparaison, c'est comme si nous voulions modéliser la circulation des voitures et la formation des embouteillages qui y est potentiellement associée sans prendre en compte la configuration des rues, des routes et des autoroutes. Un tel modèle pourrait conduire à une seule recommandation : pour contrôler les flux, tout le monde doit rester chez soi. Personne ne comprendrait qu'on puisse utiliser un tel modèle pour prédire les flux du trafic et tout le monde trouverait la recommandation inadaptée. Dans le cas des modèles SIR nous nous trouvons précisément dans cette situation : l'infrastructure routière correspond à la structure des interactions sociales qui est absente de ces modèles. Il est donc impossible d'indiquer quels acteurs et quels chemins spécifiques il vaudrait mieux emprunter pour ralentir la diffusion. Ainsi nous sommes obligés de nous tenir à une intervention généralisée : personne ne doit se déplacer. C'est une forme de modélisation sans modélisation et une façon de choisir sans choisir.

Un agenda pour la modélisation des prochaines épidémies

Le potentiel de l'intégration de la structure des interactions sociales dans la modélisation de la diffusion des pathogènes est donc énorme. Cela nous permettrait de dessiner des interventions ciblées sur les individus et les chemins de transmission qui ont le plus grand impact sur l'épidémie au lieu de demander à l'ensemble de la population de supporter des coûts inhumains. Pour rendre cette stratégie opérationnelle, il nous faut cependant des données très riches et des modèles exigeants en termes de puissance de calcul.

On peut d'ailleurs formuler l'hypothèse que, si les modèles SIR continuent de dominer la scène de la modélisation des épidémies, c'est que, en plus d'un formalisme puissant et parcimonieux (les équations différentielles, le plus souvent), ils demandent des données empiriques moins riches. Les premières critiques explicites de la rigidité de cette approche pour modéliser l'hétérogénéité des comportements individuels ainsi que la structure des interactions sociales dans le cadre des épidémies remontent en effet au moins à la pandémie du virus H5N1. À ce moment-là, l'idée d'utiliser des approches algorithmiques, notamment la simulation à base d'agents artificiels, avait été puissamment mise en avant par Joshua Epstein (2009)⁴. Or, comme cela a été souligné avec force dans un manifeste récemment publié dans le *Journal of Artificial Societies and Social Simulation* (<http://jasss.soc.surrey.ac.uk/23/2/10.html>), ces modèles sont toujours largement absents du débat public sur la diffusion des virus. C'est que, faute de données sur les interactions sociales suffisamment détaillées et d'infrastructures informatiques adaptées, nous sommes toujours dans l'impossibilité de mettre en œuvre ces outils à grande échelle.

Aussi dramatique soit-elle, la situation actuelle devrait donc nous faire prendre conscience de l'ignorance qui est la nôtre. Elle devrait pousser la puissance publique à réaliser que les sciences sociales quantitatives ont besoin de l'équivalent d'un accélérateur de particules ou d'un centre national pour la prévision météorologique. Tout le monde comprend aisément que pour avancer, par exemple, dans la compréhension de la structure moléculaire d'un nouveau virus, il faille des outils de calcul d'une puissance inouïe. Qui serait prêt en revanche à soutenir des efforts comparables, sur le long terme, pour arriver à modéliser et simuler la diffusion du même virus porté par des chaînes d'interactions humaines aussi complexes, si ce n'est pas plus, que les interactions chimiques et cellulaires au sein du virus ? Aussi insensée que cette idée puisse paraître, la situation dans laquelle nous nous trouvons devrait nous faire comprendre que, sans pouvoir modéliser et simuler les comportements hétérogènes au sein de structures d'interactions complexes à l'échelle d'un pays, nous ne pouvons simplement pas intervenir de manière plus ciblée que les modèles SIR, même dans leurs variantes individu-centrées, ne le permettent. Pour avancer dans cette direction, on peut indiquer au moins quatre chantiers où des efforts massifs devraient être consentis.

1/ Il faut poursuivre la collecte à grande échelle de différents types de données digitales telles que les flux d'appels téléphoniques, les échanges sur les réseaux virtuels fondés sur l'Internet ou encore les données concernant différentes formes de mobilité dans l'espace. Ces données ont un grand potentiel, car il a été démontré qu'elles peuvent être utilisées pour inférer la présence d'interactions entre les acteurs.

2/ Il ne faut pas cependant délaisser la description la plus directe possible des interactions que les acteurs réalisent effectivement en face à face au fil de leurs activités. Ceci est particulièrement évident quand il s'agit d'étudier la transmission d'un virus. Alors que nous recensons une très grande quantité de données sur les propriétés des individus, des foyers ou

⁴ Joshua M. Epstein, « Modelling to contain pandemics », *Nature*, 460, 7256, (2009), 687 doi: 10.1038/460687a.

des entreprises singulièrement considérées, une cartographie des réseaux sociaux fait toujours défaut à notre pays. Pour ce faire, il faut imaginer des manières nouvelles de combiner des dispositifs numériques innovants (tel que ceux utilisés dans le projet *SocioPatterns*, <http://www.sociopatterns.org/>), notamment pour capturer l'évolution au fil du temps des interactions entre les acteurs, avec des outils plus anciens, notamment l'échantillonnage aléatoire (qui a fait des progrès considérables en matière de réseaux sociaux dans les dernières années).

3/ il faut poursuivre le développement de méthodes pour traiter les données sur les interactions sociales. Leur volume, leur multiplicité et leurs évolutions dans le temps sont tels que des méthodes adaptées doivent être développées pour traiter ces informations ainsi que pour corriger les erreurs et les omissions qui peuvent en miner la fiabilité. Il faut également continuer à développer les techniques qui existent pour reconstituer le réseau global d'un type d'interactions donné à partir de données (lacunaires) sur des réseaux égo-centrés.

4/ il faut enfin renforcer les méthodes qui permettent d'étudier les conséquences macroscopiques des données sur la structure des interactions, en particulier quand la structure de ces interactions se combine avec des comportements humains d'un type plutôt que d'un autre. Les avis du conseil scientifique ayant informé les décideurs politiques, ainsi que les publications des modélisateurs auxquelles ces avis se réfèrent, insistent sans cesse, mais verbalement, sur le fait que le résultat des mesures prises dépend du comportement des acteurs. Pourtant, aucune des modélisations formelles supportant le processus décisionnel en cours ne représente explicitement ces comportements en produisant des *scénarios* alternatifs en fonction de ces comportements (comme les gestes barrière, par exemple). Les mêmes techniques de modélisation et de simulation informatique qui sont en principe capable d'intégrer des informations fines sur la structure des interactions sociales, à savoir les modèles algorithmiques à base d'agents artificiels, peuvent également incorporer une représentation aussi détaillée qu'on le souhaite des comportements (eux aussi hétérogènes) des acteurs.

Pour chacun de ces développements, il y a évidemment un problème de coût pour déployer les différents dispositifs de recherche à grande échelle. La gravité de la situation impose à la politique scientifique et la puissance publique de se saisir de ces défis. Le caractère ciblé des interventions pour altérer un processus de diffusion est proportionnelle à la finesse des informations dont nous disposons sur les acteurs et leurs interactions ainsi qu'à la puissance et au réalisme des modèles dont nous disposons pour traiter ces informations. S'engager résolument sur cette voie peut évidemment se heurter aux réticences que chacun d'entre nous peut avoir à accepter de rendre visible une partie de sa vie relationnelle, même sous des procédures encadrées et contrôlées. À ce sujet, il est utile de rappeler la réponse de Renaud Piarroux au journaliste lui demandant ce que, en tant que médecin, il pourrait faire avec des informations sur les porteurs et leurs contacts si ces informations étaient anonymisées : « cela transforme les cas en points et je ne sais pas faire porter un masque à un point, ni l'aider à protéger ses proches » (*Le Monde*, 4 Avril 2020, p. 4). Le choix pour le futur est donc clair. Ne

pas consentir à cette petite contribution à la vie collective nous expose à nous contraindre toutes et tous à des privations de libertés infiniment plus lourdes.

Références

- Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José J. Ramasco, Stefano Merler, Alessandro Vespignani, « Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models », *BMC infectious diseases*, vol. 10, n° 190, 2010 (doi: 10.1186/1471-2334-10-190).
- Albert-Laszlo Barabasi, *Linked*, New York, Basics Books, 2014.
- Albert-Laszlo Barabási, Eric Bonabeau, « Scale-Free Networks », *Scientific American*, vol. 288, 2003, p. 60-69.
- Alain Barrat, Marc Barthélemy, Alessandro Vespignani, *Dynamical Processes on Complex Networks*, Cambridge, Cambridge University Press, 2008.
- Anton Camacho, Adam J. Kucharski, Rachel Lowe, Rosalind M. Eggo, W. John Edmunds, « Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15 » *PLoS Computational Biology*, vol. 15, n° 2, e1006785. (<https://doi.org/10.1371/journal.pcbi.1006785>).
- Laura Di Domenico, Giulia Pullano, Chiara E. Sabbatini, Pierre-Yves Boëlle, Vittoria Colizza, « Expected impact of lockdown in Île-de-France and possible exit strategies », Report #9 (<https://www.epicx-lab.com/covid-19.html>).
- Wei Duan , Zongchen Fan, Peng Zhang, Gang Guo, Xiaogang Qiu, « Mathematical and computational approaches to epidemic modeling: a comprehensive review » *Frontiers in Computational Science*, vol. 9, n° 5, 2015, p. 806-826. DOI 10.1007/s11704-014-3369-2.
- Fredrik Liljeros, Christofer R. Edling, Luís A. Nunes Amaral, H. Eugene Stanley, Yvonne Aberg « The web of human sexual contacts » *Nature*, vol. 411, n° 6840, 2001, p. 907-8.
- Matthew O. Jackson, Dunia Lopez-Pintado, « Diffusion and contagion in networks with heterogeneous agents and homophily » *Network Science*, vol. 1, n° 1, 2013, p. 49-67. (<https://www.cambridge.org/core/search?filters%5BauthorTerms%5D=DUNIA%20L%20PEZ-PINTADO&eventCode=SE-AU>).
- Matt Keeling, Pejman Rohani, *Modeling Infectious Diseases in Humans and Animals* Princeton, Princeton University Press, 2008.

- Stephen Kissler, Christine Tedijanto, Marc Lipsitch, Yonatan H. Grad, « Social distancing strategies for curbing the COVID-19 epidemic 2020 » (<https://www.medrxiv.org/content/10.1101/2020.03.22.20041079v1.full.pdf>).
- James Moody, Richard A Benton, « Interdependent effects of cohesion and concurrency for epidemic potential », *Annals of epidemiology* vol. 26, n° 4, 2016, p. 241-8. doi:10.1016/j.annepidem.2016.02.011
- Martina Morris, Mirjam Kretzschmar, « Concurrent partnerships and the spread of HIV », *AIDS*, vol.11, 1997, p. 641-648.
- Luis E. C. Rocha, Fredrik Liljeros, Petter Holme, « Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts », *PLoS Computational Biology*, vol. 7, n° 3, 2011, e1001109. doi:10.1371/journal.pcbi.1001109.
- Duncan J. Watts, Steven H. Strogatz, (1998), « Collective dynamics of ‘Small-world’ networks », *Nature*, vol. 393, 1998, p. 440-442.

Publié dans lavedesidees.fr, le 21 avril 2020.