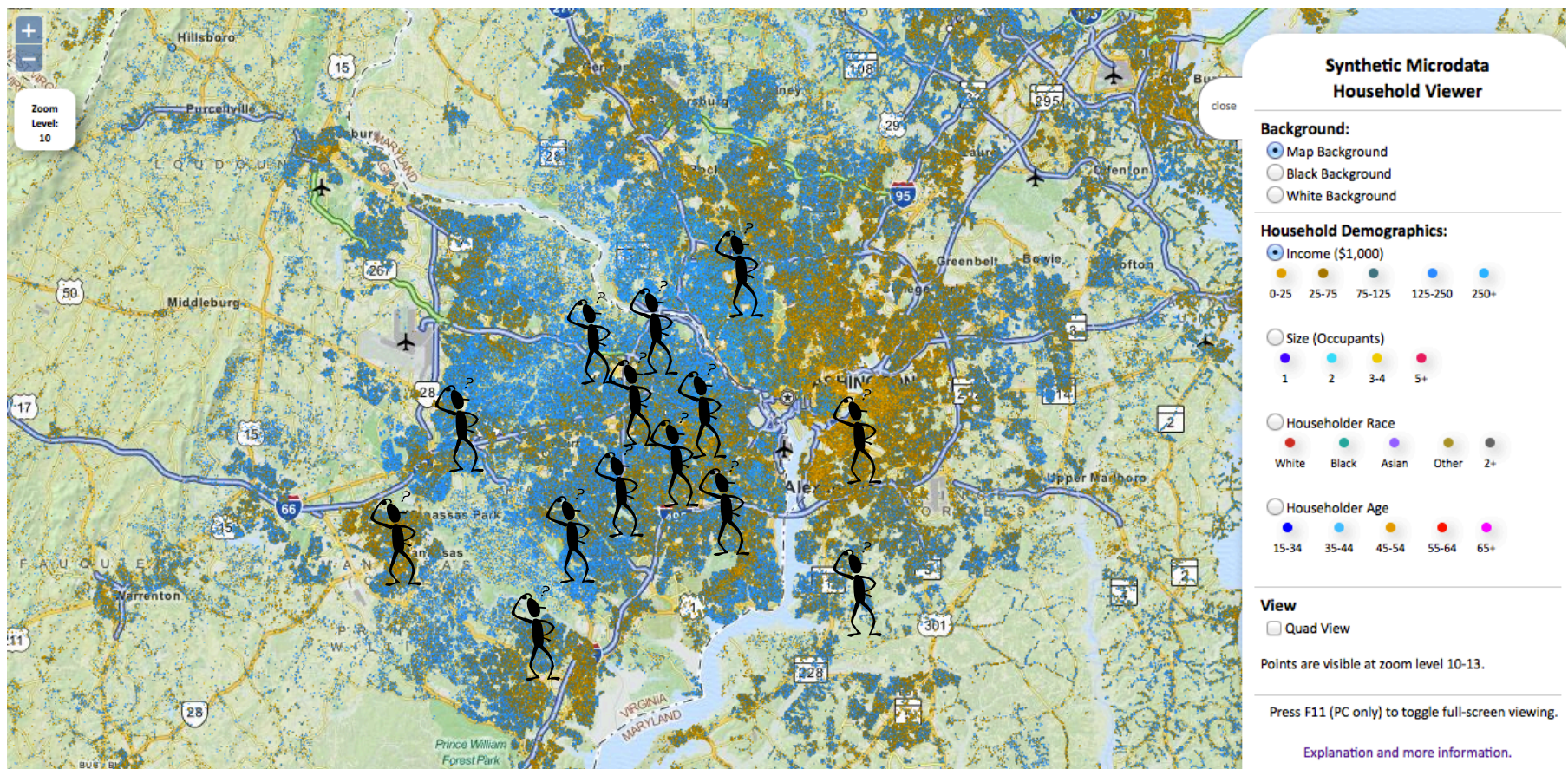




Génération de Populations Synthétiques pour la Simulation Sociale

Frédéric AMBLARD

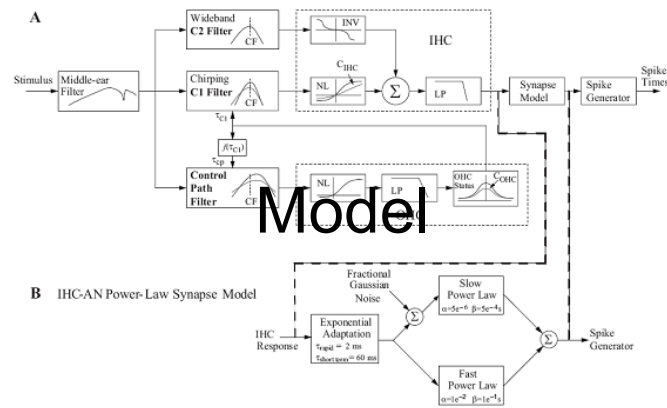


Simulation Sociale



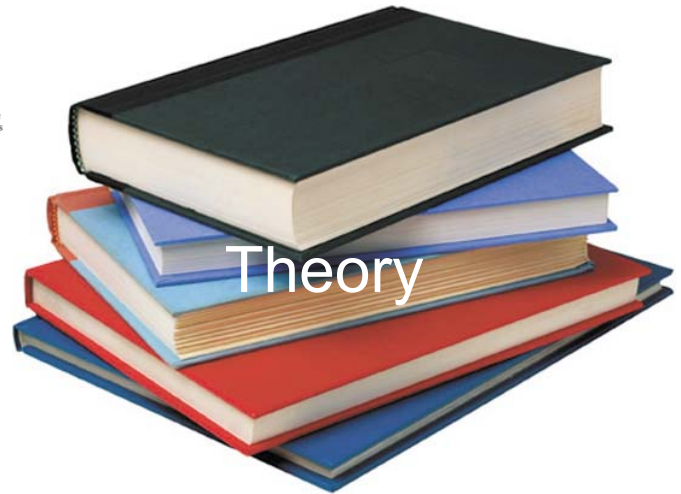
Modelling as a Third way of doing Social Science

Data



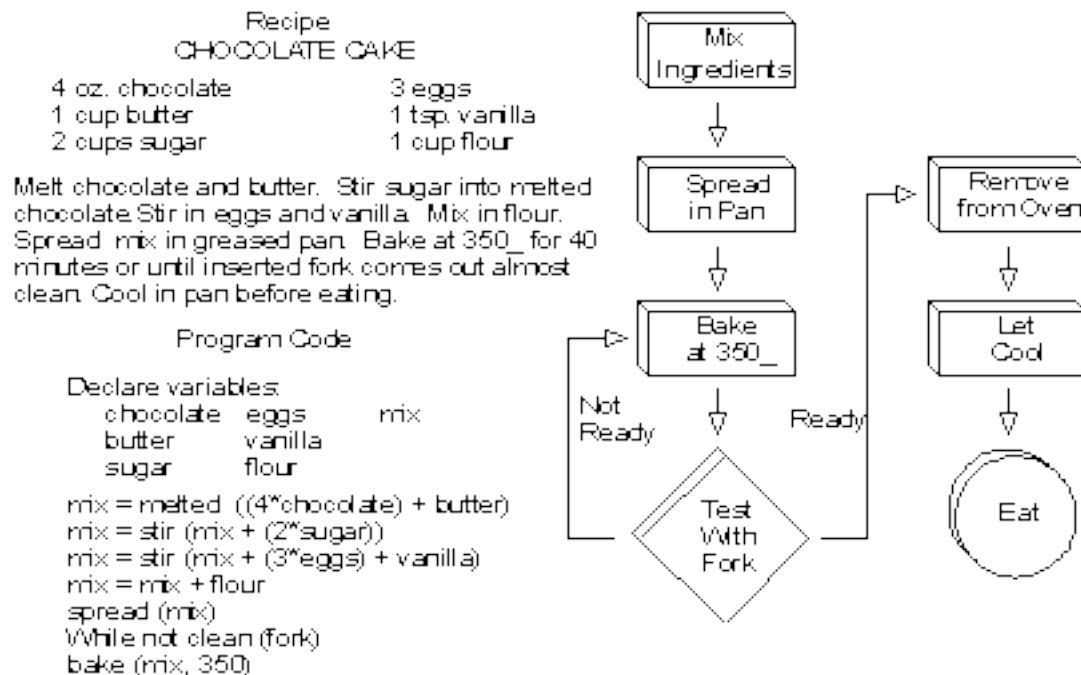
Theory

Experiments



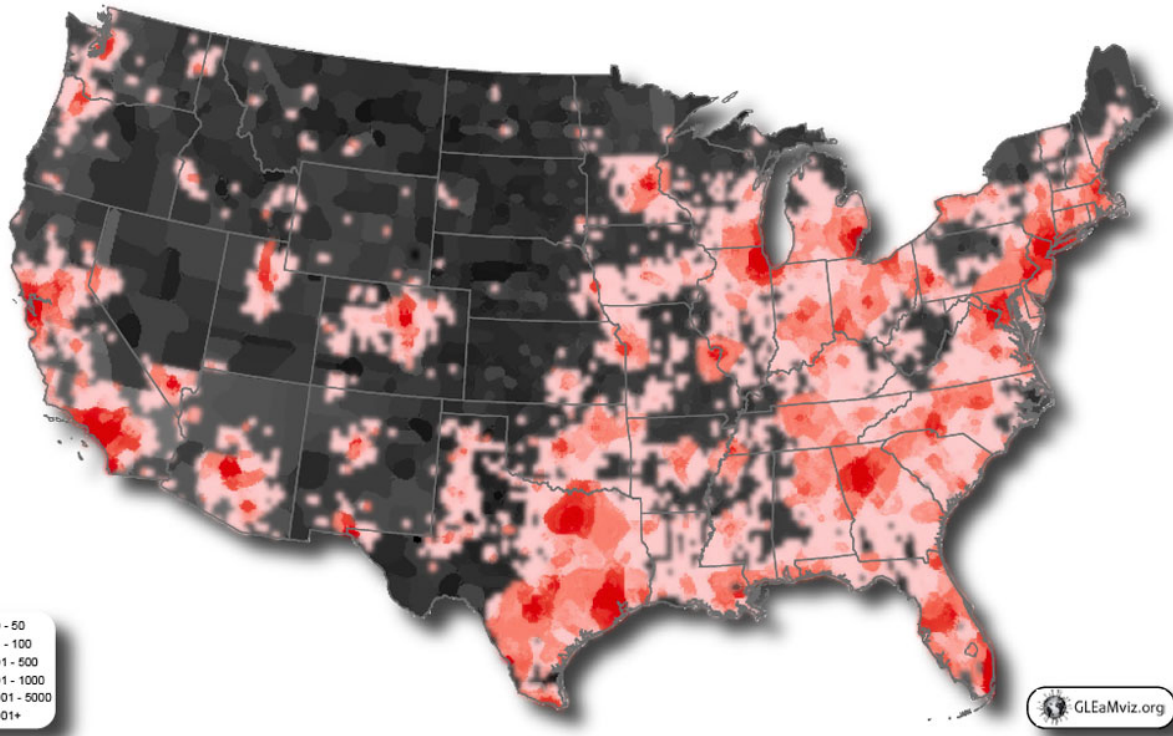
What Computer Science could propose?

➤ Algorithms



What Computer Science could propose?

➤ Simulation



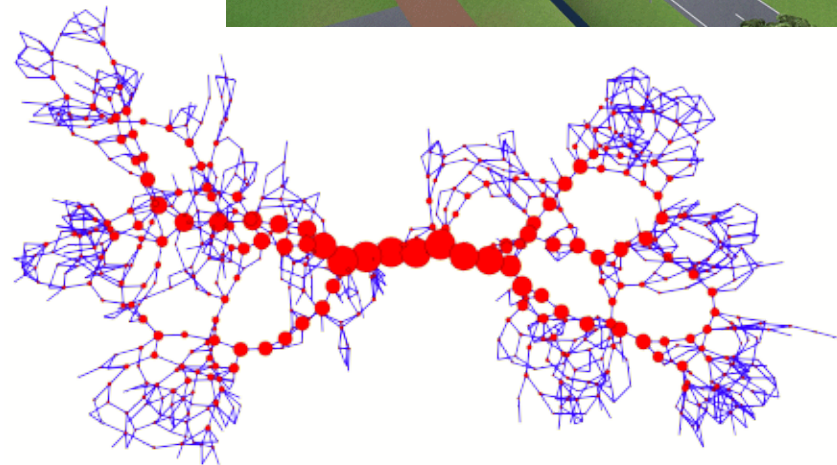
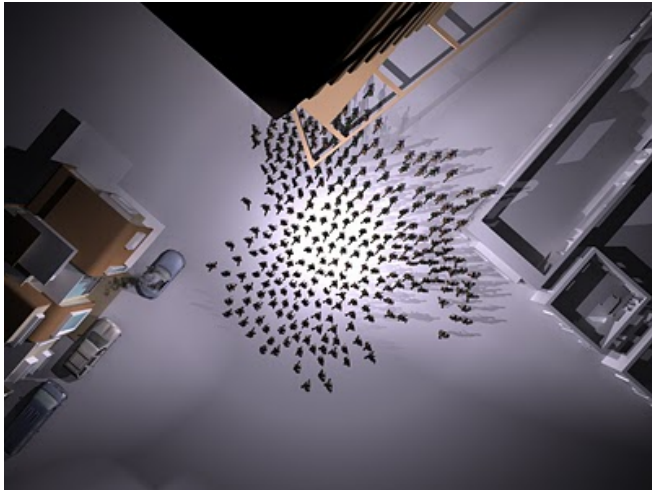
What Computer Science could propose?

➤ Individual-based approach

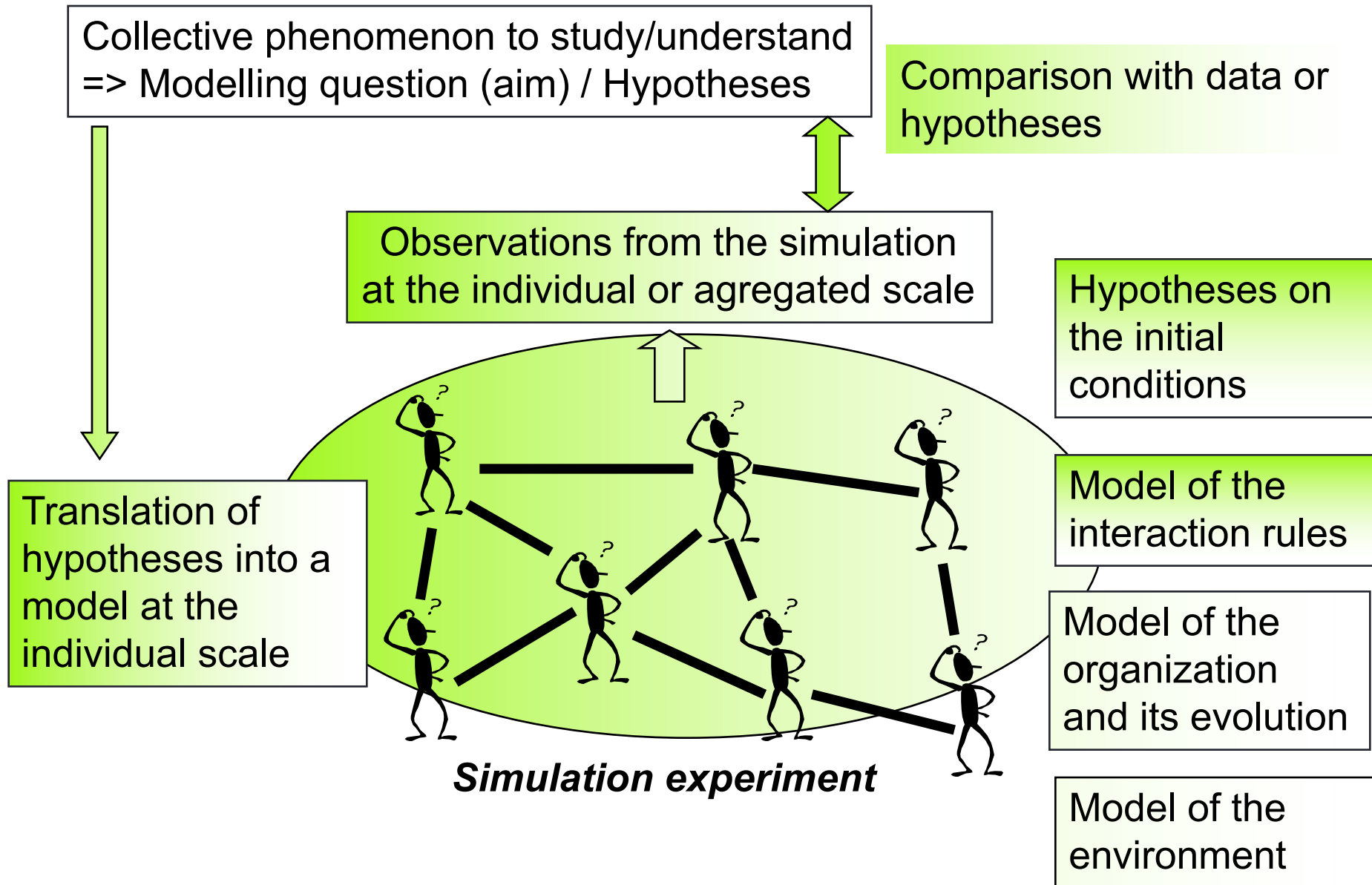


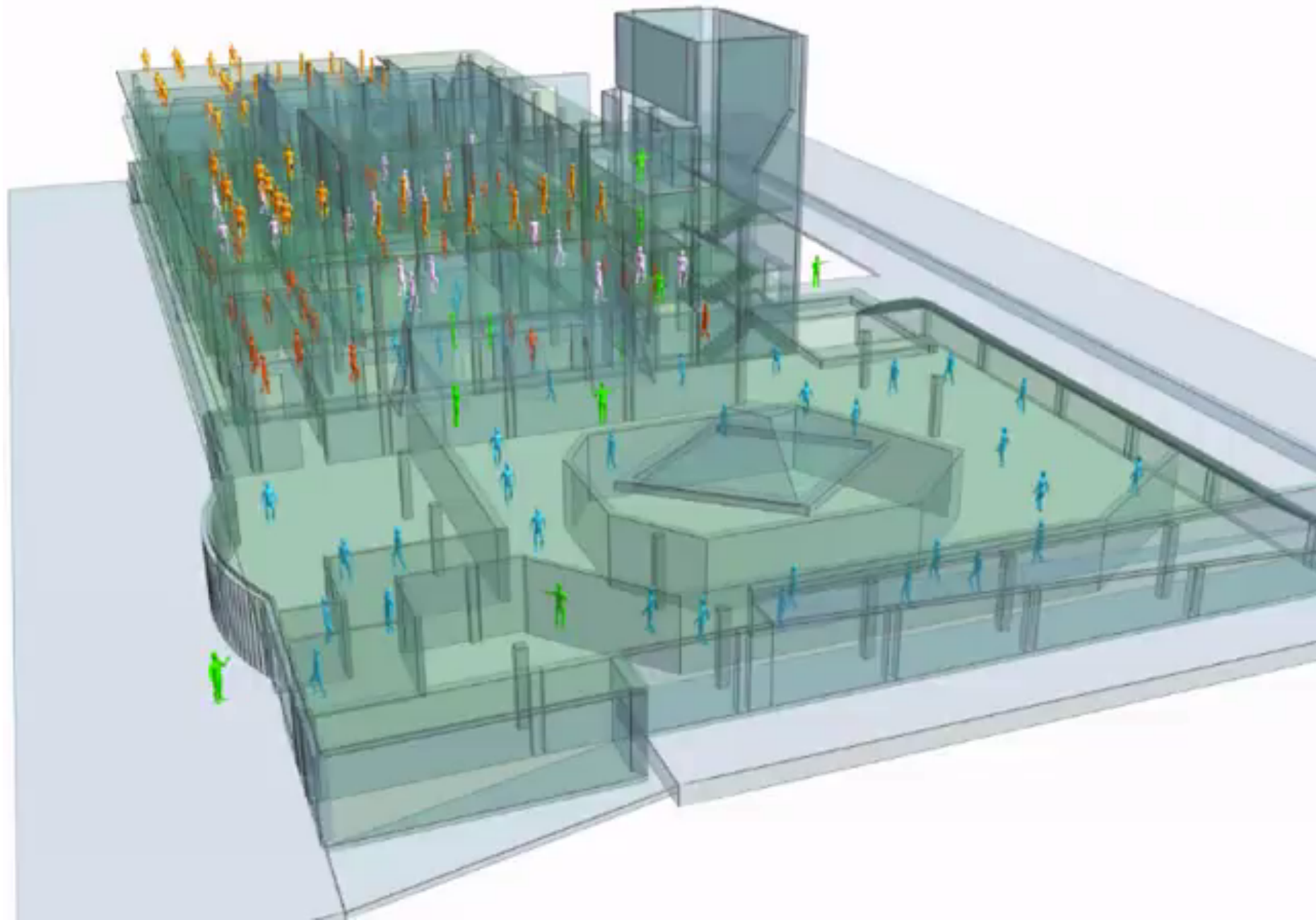
What Computer Science could propose?

- Taken together you obtain agent-based social simulation



Agent-Based Social Simulation





Some remarks concerning Agent-Based Modelling

➤ Advantages

- Hypotheses expressed at the individual level
- Modelling the dynamics
- Models are experimental objects (simulation)
- Many different possible use

➤ Drawbacks

- Reproduction of the complexity of the real system (micro/macro link)
- Difficulty to understand how the results are produced
- Validation

Génération de populations synthétiques

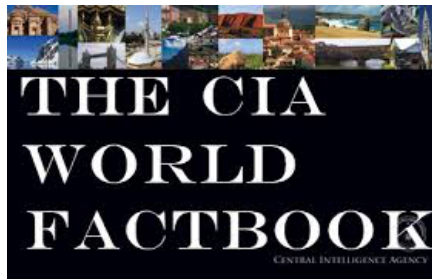


Pourquoi générer des populations synthétiques ?

- Initialiser ma simulation
- Introduire de l'hétérogénéité dans les modèles
- Réalisme des scénarios testés
 - Ex: Diffusion d'épidémie, tester des stratégies de fermeture d'école, de vaccination, de mise en quarantaine de populations cibles...



A partir de quoi ?



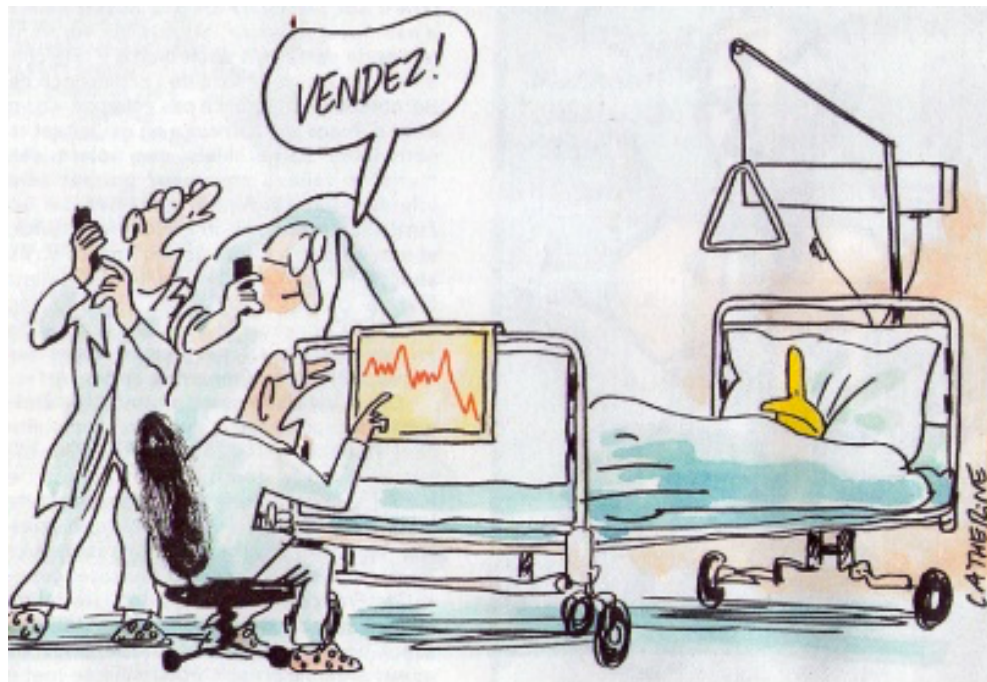
Problèmes sous-jacents

- Granularités des données
 - Données agrégées
 - Données individuels
 - Sampling / Profils types



Problèmes sous-jacents

- Granularités des données
- Qualité des données



Problèmes sous-jacents

- Granularités des données
- Qualité des données
- Croisement de données
 - Granularité / temporalité : croisabilité des données
 - Secret statistique



Problèmes sous-jacents

- Granularités des données
- Qualité des données
- Croisement de données
- Gagner en généricité vs. Approches ad hoc
 - <http://www.genstar.fr>



Générer des populations synthétiques

- Trois dimensions
 - Génération individus



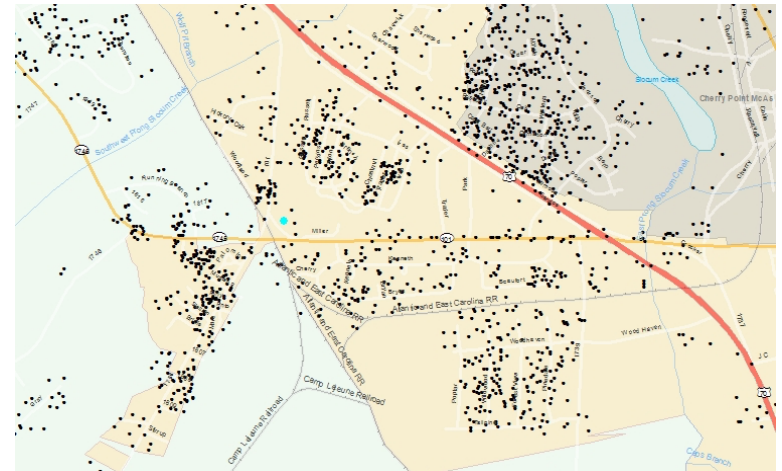
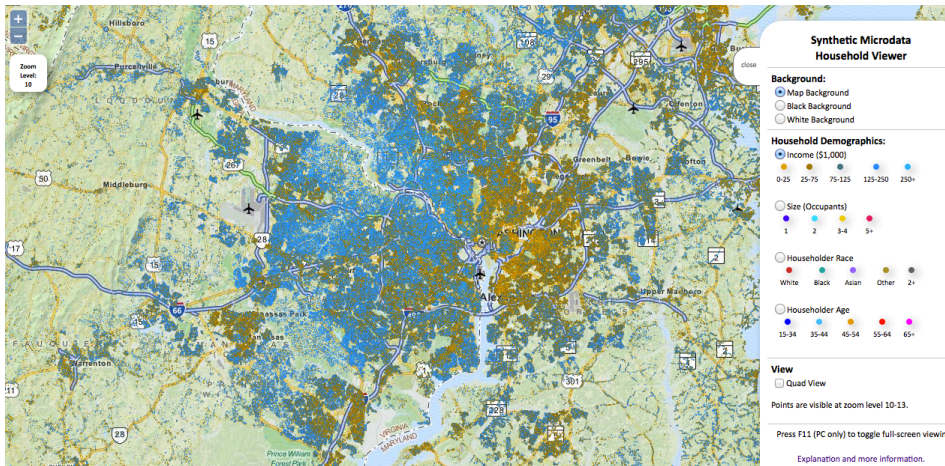
Générer des populations synthétiques

- Trois dimensions
 - Génération individus
 - Structuration sociale : création de foyers



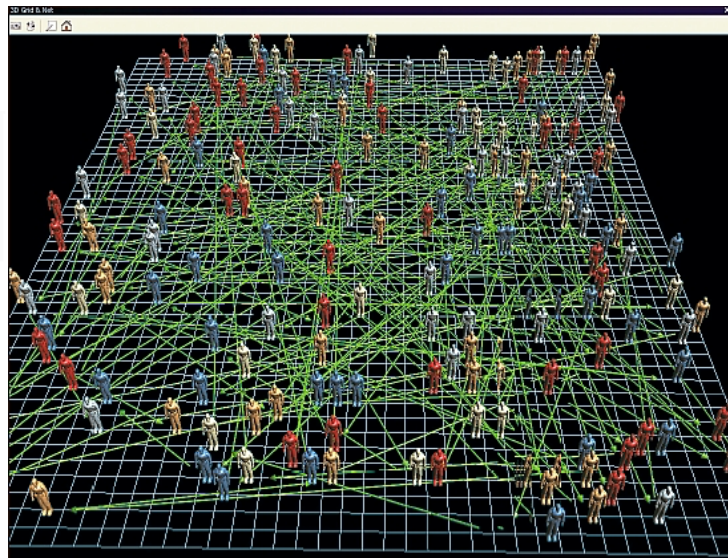
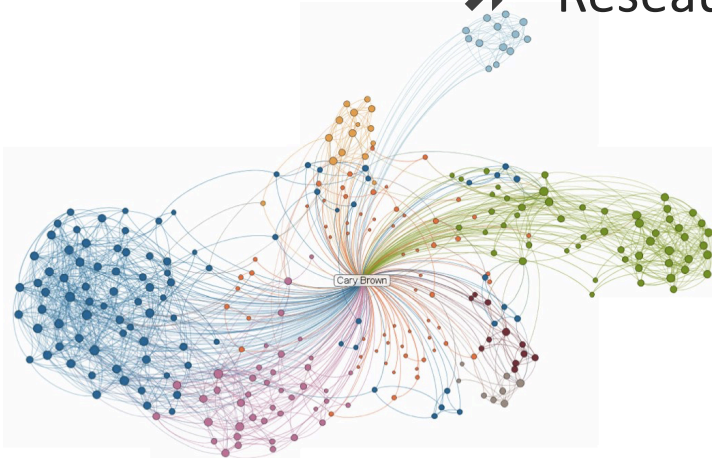
Générer des populations synthétiques

- Trois dimensions
 - Génération individus
 - Structuration sociale : création de foyers
 - Localisation des individus/foyers



Générer des populations synthétiques

- Trois dimensions
 - Génération individus
 - Structuration sociale : création de foyers
 - Localisation des individus/foyers
 - Réseaux sociaux synthétiques

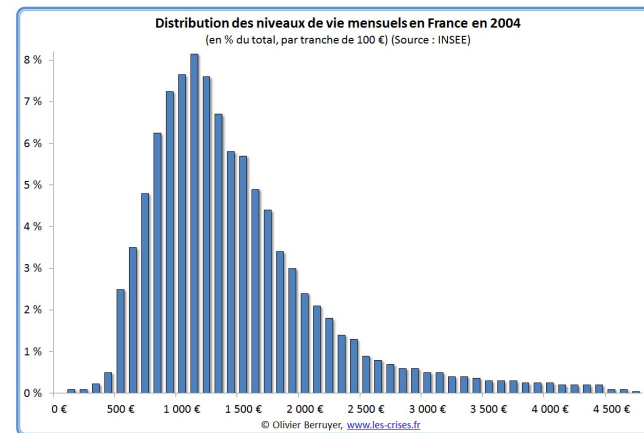
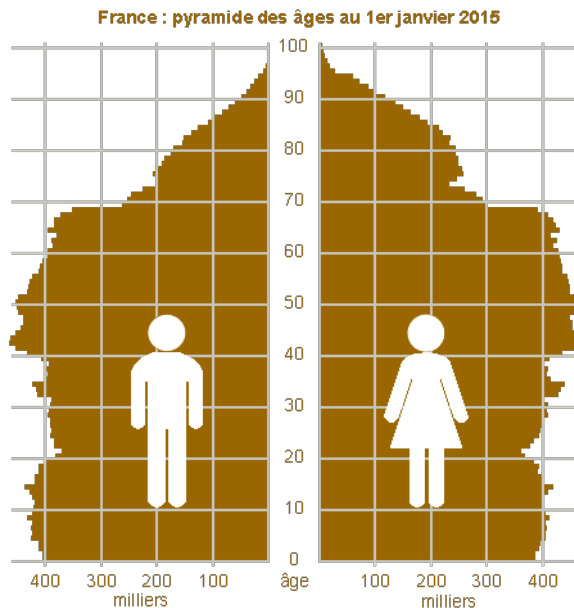


➤ Génération individus

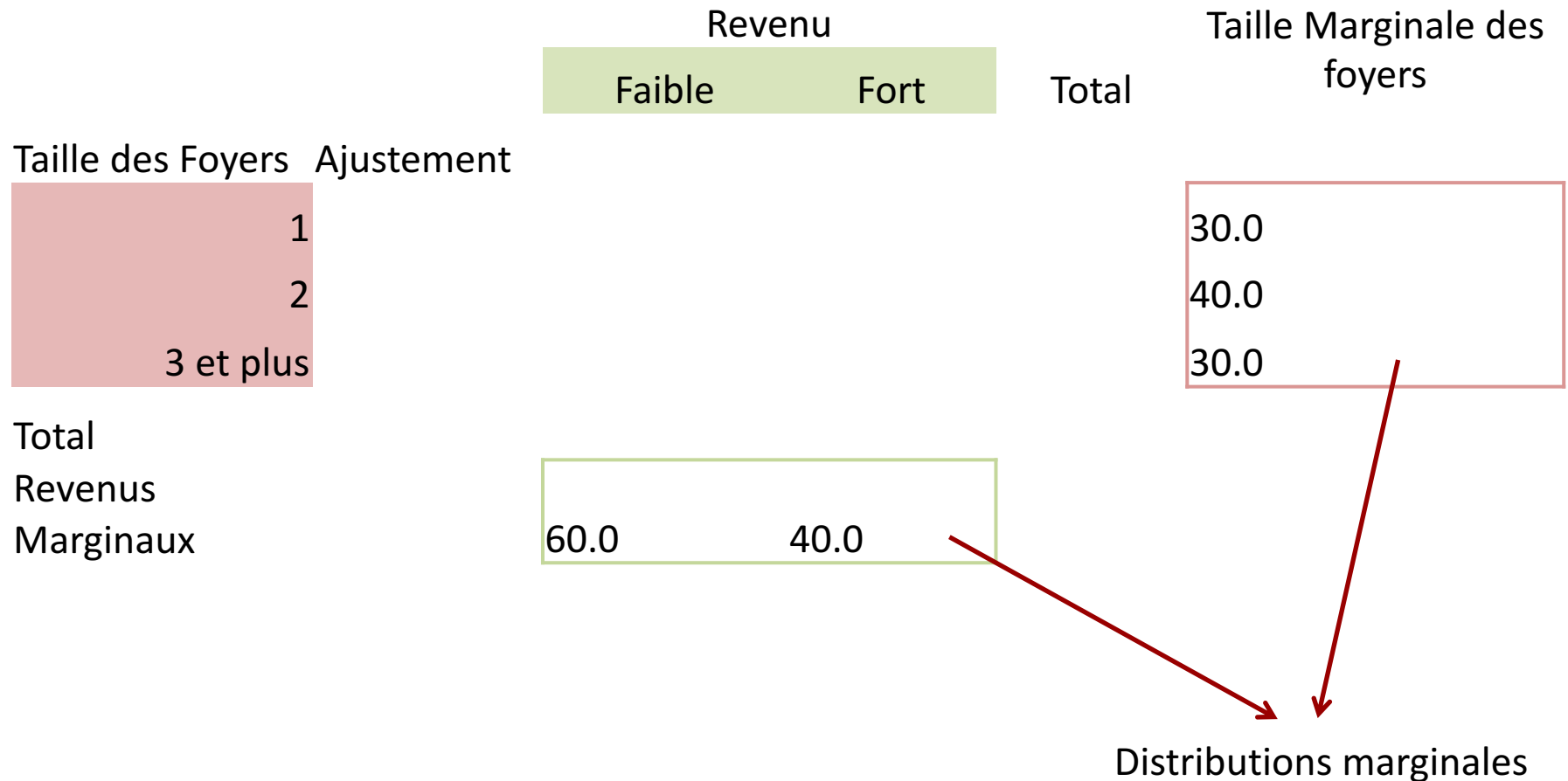
- Ex: pyramide âges, distribution revenus
- Sol simple: tirage aléatoire indépendants
- Pb: ce n'est pas indépendant

➤ Tables de contingence

➤ Introduction ad hoc de règles (modélisation)



Iterative Proportional Fitting (IPF)



Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement				
	1	3.0	1.0	4.0	30.0
	2	2.0	4.0	6.0	40.0
	3 et plus	2.0	1.0	3.0	30.0
Total		7.0	6.0		
Revenus					
Marginaux		60.0	40.0		

Données initiales
Echantillon (graine)

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	60/7 = 8.57	40/6 = 6.66		
	1	3.0	1.0	4.0	30.0
	2	2.0	4.0	6.0	40.0
	3 et plus	2.0	1.0	3.0	30.0
Total		7.0	6.0		
Revenus					
Marginaux		60.0	40.0		

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	8.57	6.66		
		$3 * 8.57 =$			
	1	25.7	$= 1 * 6.66$	4.0	30.0
	2			6.0	40.0
	3 et plus			3.0	30.0
Total		7.0	6.0		
Revenus					
Marginaux		60.0	40.0		

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	8.57	6.66		
	1	25.7	6.7	32.4	30.0
	2	17.1	26.7	43.8	40.0
	3 et plus	17.1	6.7	23.8	30.0
Total		60.0	40.0		
Revenus					
Marginaux		60.0	40.0		

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	8.57	6.66		
	1 = $30/32.4$	25.7	6.7	32.4	30.0
	2	17.1	26.7	43.8	40.0
	3 et plus	17.1	6.7	23.8	30.0
Total		60.0	40.0		
Revenus					
Marginaux		60.0	40.0		

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	8.57	6.66		
	1 0.93	=25.7*0.93	6.7	32.4	30.0
	2 0.91	17.1	26.7	43.8	40.0
3 et plus	1.26	17.1	6.7	23.8	30.0
Total		60.0	40.0		
Revenus					
Marginaux		60.0	40.0		

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	8.57	6.66		
	10.93	23.8	6.2	30.0	30.0
	20.91	15.7	24.3	40.0	40.0
3 et plus	1.26	21.6	8.4	30.0	30.0
Total		61.1	38.9		
Revenus					
Marginaux		60.0	40.0		

Iterative Proportional Fitting (IPF)

		Revenu		Total	Taille Marginale des foyers
		Faible	Fort		
Taille des Foyers	Ajustement	8.57	6.66		
1	1.00	23.6	6.4	30.0	30.0
2	1.00	15.2	24.8	40.0	40.0
3 et plus	1.00	21.3	8.7	30.0	30.0
Total		60.0	40.0		
Revenus					
Marginaux		60.0	40.0		

Après 3 itérations ...

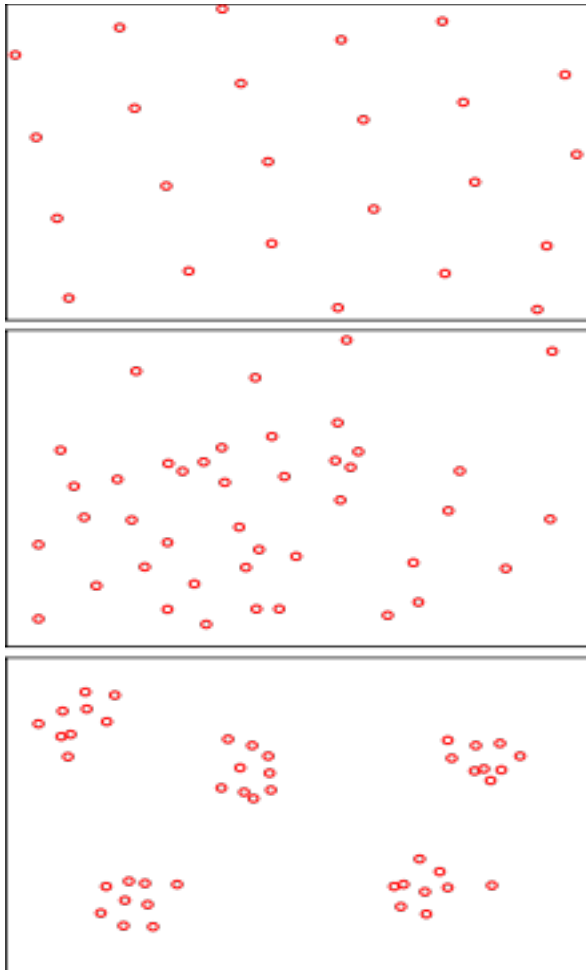
Iterative Proportional Fitting (IPF)

- Utilisable à n dimensions
- A différentes étapes de la génération
 - Génération d'individus
 - Génération des ménages
 - Génération des attributs spatiaux

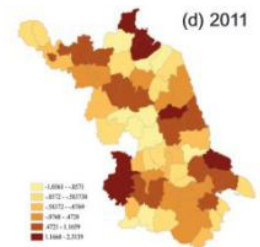
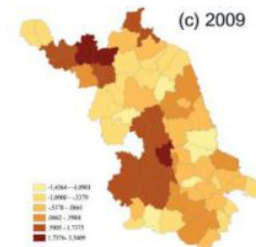
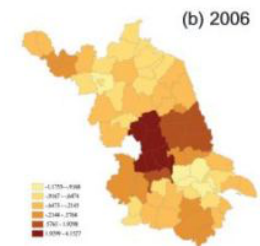
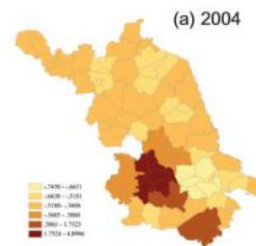
Méthodologie générale pour la génération de population

- Génération des individus
- Génération des ménages
- Allocation des individus aux ménages
- Génération des entités spatiales
- Allocation des ménages aux entités spatiales
- ...

Localisation des individus



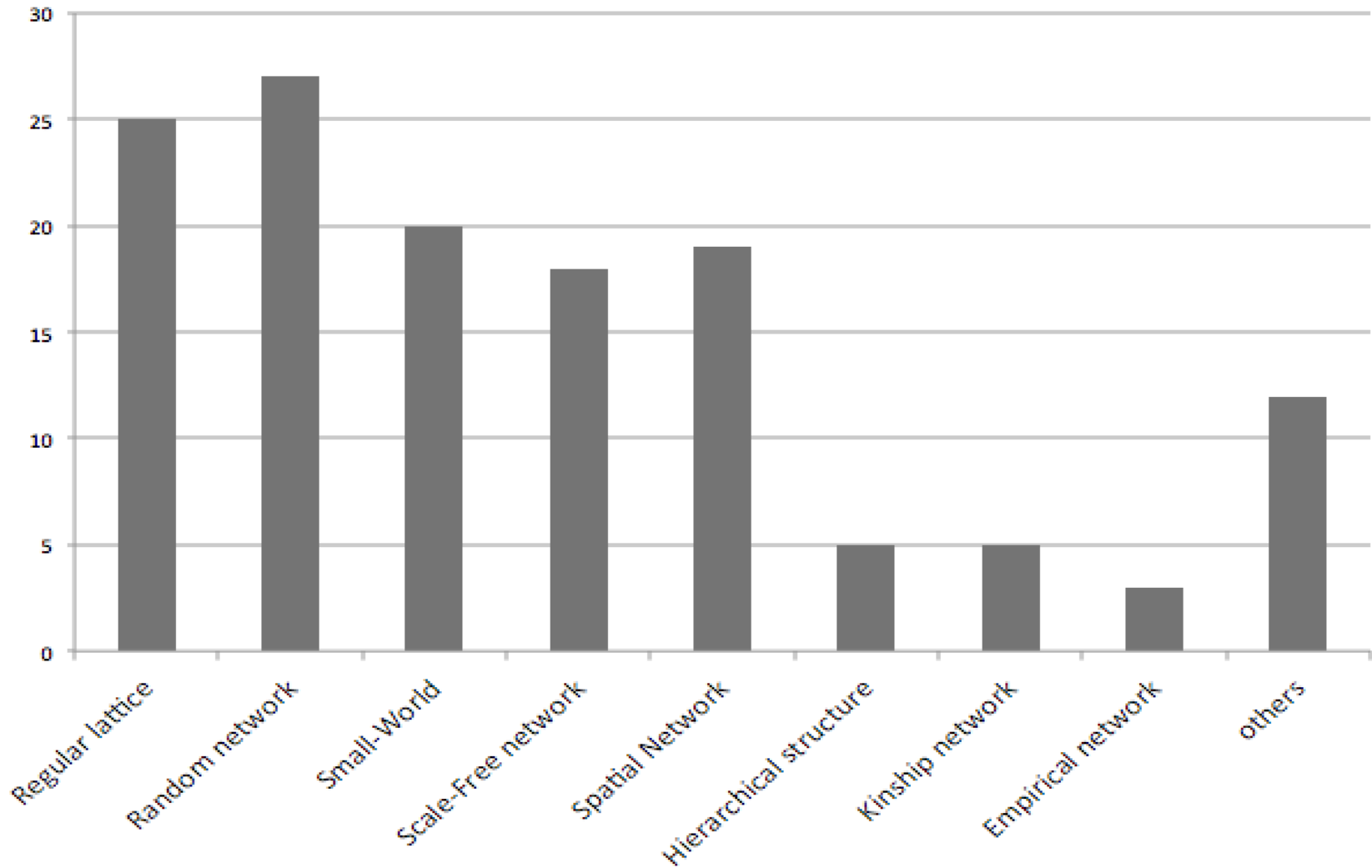
- Affectation des individus à des parcelles
- Distributions spatiales (semis de points)
- Approches hybrides



Génération de réseaux sociaux synthétiques

- Problèmes de données disponibles
- Mais ... in fine simulation sur réseaux (ex: épidémio)
 - Modèles abstraits
 - Scale-Free Networks
 - Small-Worlds
 - Random Networks (Erdős-Rényi)

Pratiques en simulation sociale (JASSS)



Génération de réseaux sociaux synthétiques

- Problèmes de données disponibles
- Mais ... in fine simulation sur réseaux (ex: épidémio)
 - Modèles abstraits (SFN, SW, RN) cf JASSS
 - Modèles par échantillonnage (ERGM, ...) mais il faut des échantillons
 - Proposition Roth et Cointet réutilisation de données empiriques issues d'un autre contexte
 - Construction de réseau à base de règles YANG (Samuel Thiriot)

Approche mémétique

- Thèse Audren Bouadjio-Boulic
- Partir de règles sociales (les amis de mes amis sont mes amis)
- Les règles se diffusent sur le réseau social et le modifient
- Avantage: approche locale qui permet de s'affranchir des problèmes d'échelle
- Inconvénient: problèmes de maîtrise de la solution (calibration)

References

- Synthetic Population Generation Without a Sample. Johan Barthelemy, Philippe L. Toint. *Transportation Science, Articles in Advance*, pp. 1–14, 2012.
- Barrett, C. L., Beckman, R. J., Khan, M., Anil Kumar, V. S., Marathe, M. V., Stretz, P. E., ... & Lewis, B. (2009, December). Generation and analysis of large synthetic social contact networks. In *Winter Simulation Conference* (pp. 1003-1014). Winter Simulation Conference.
- Smith, D. M., Clarke, G. P., & Harland, K. (2009). Improving the synthetic data generation process in spatial microsimulation models. *Environment and planning. A*, 41(5), 1251.
- Gargiulo, F., Ternes, S., Huet, S., & Deffuant, G. (2010). An iterative approach for generating statistically realistic populations of households. *PloS one*, 5(1), e8828.
- Luotsinen, L. J. (2014, March). POPSIM: a platform targeting the modeling and simulation of human populations in urban environments. In *Proceedings of the 7th International ICST Conference on Simulation Tools and Techniques* (pp. 160-165). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).