HiDALGO

Model Reduction in HiDALGO – Initial Plans and Ideas

B. C. Geiger[1], C. Schweimer[1], D. Suleimenova[2], D. Groen[2], C. Gfrerer[3], D. Pape[3], R. Elsässer[3], T. A. Kocsis[4], Z. Horváth[4]

[1]Know-Center GmbH, Graz, Austria; [2]Department of Computer Science, Brunel University, London, UK;
[3]Efficient Algorithms Group, Paris Lodron University Salzburg, Austria;
[4]Department of Mathematics and Computational Sciences, Széchenyi István University Győr, Hungary

HiDALGO Overview

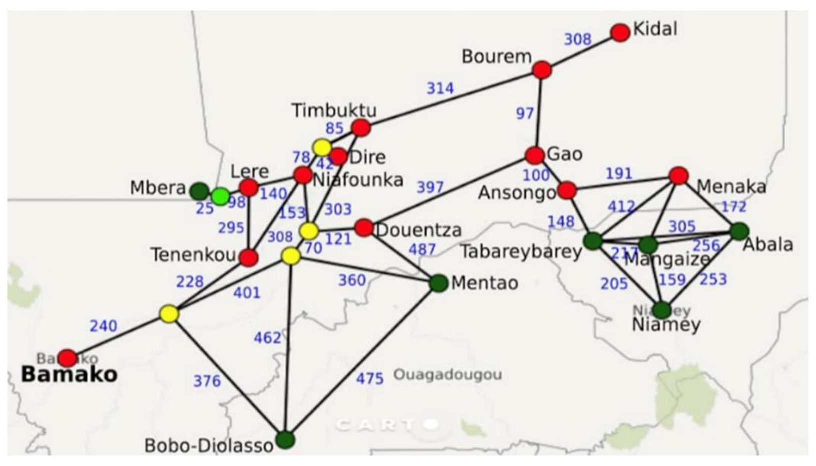
- HPC and Big Data Technologies for **Global Challenges (GC)**
- Provide approaches to solve three major GC (Air Pollution, Migration and Information Spreading in Social Networks) with Big Data and AI
- Coupled simulations based on HPC, HPDA and Cloud Computing

Urban Air Pollution Use Case

- Over 4 million premature deaths worldwide due to bad air quality[1]
- Simulation of **Air Pollution Dispersion** with Computational Fluid Dynamics (CFD) based models and real-world sensor data from urban areas
- Software framework for urban air quality prediction and control: *3DAirQC*[2]

Number of refugees as predicted by our simulation and obtained from the UNHCR data for the Mali conflict

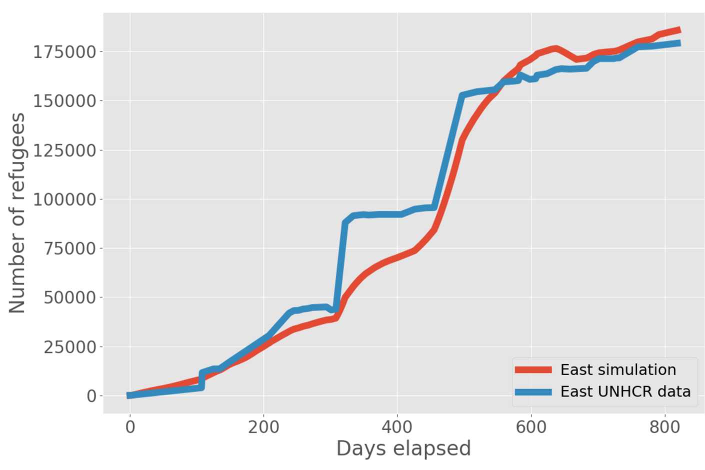
Overview of geographic network model for Mali



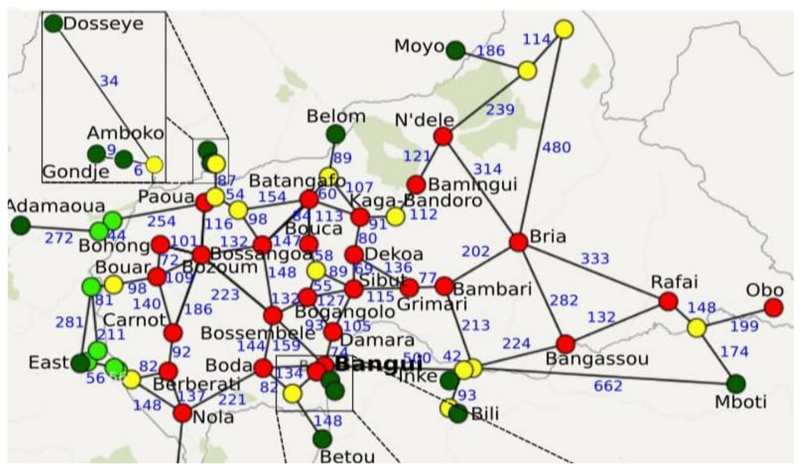
Migration Use Case

- More than 20 million refugees worldwide in recent years[3]
- Agent-based models (ABMs) to perform **refugee movement simulations**
- Accuracy of refugee destination prediction ~ 75% in three African conflicts[3,4]
- Example: Conflict in the Central African Republic (CAR) in 2013-2016[3]

Number of refugees predicted by simulation and obtained from the UNHCR data for the East Congo camp during the CAR conflict



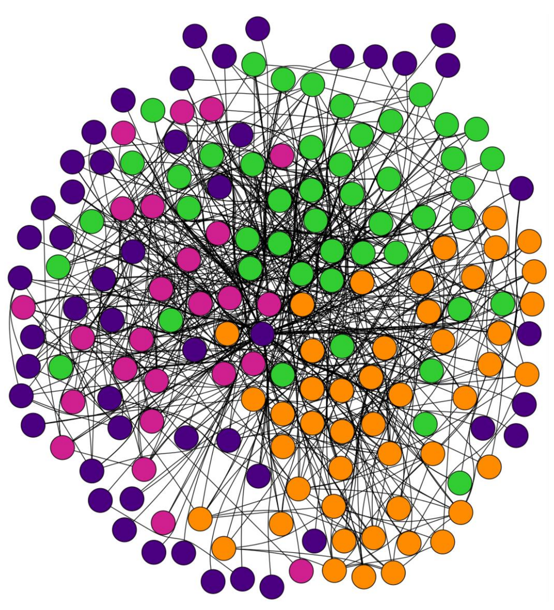
Geographic network model for CAR: conflict zones (red), camps (dark green), hubs (light green), other settlements (yellow)



Social Networks Use Case

- About 500 million daily tweets influence human/social behaviour
- Determine **characteristics of social networks** such as degree distribution, distance distribution, and community structure
- Understand, model and simulate the **spread of messages** in social networks

An example of a social network with four communities



Acknowledgments

The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

Brunel University: This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement nos. 800925 and 671564.

Model Reduction in HiDALGO

POD for CFD Simulations

- CFD Simulations based on Nonlinear Partial Differential Equation (PDE)
- Replace a system of high-dimensional PDEs $(\dim = n)$ with a system of lower-dimensional PDEs via *Proper Orthogonal Decomposition* (POD):
  - Set of snapshots $\{y_1, \dots, y_m\} \in \mathbb{R}^n$ at times $t_1, \dots, t_m$ with $\Delta t = t_i - t_{i-1}$
  - Matrix $A = [y_1 \dots y_m] \in \mathbb{R}^{n \times m}$
  - POD: $A = U \cdot \Sigma \cdot V^T$ with U, V^T orthogonal and Σ diagonal → singular value decomposition
  - First k singular vectors $\varphi_1, \dots, \varphi_k \in U,\ k \le m$ → POD basis
- Next Steps/Issues:
  - POD is inaccurate for high Reynolds number turbulent flows
  - PDEs with bifurcating solutions and subdivided parameter domains:
    - Single global POD basis often unreliable
    - Construct several local POD bases for each snapshot cluster corresponding to a different part of the parameter domain[5]

Lumpability for Agent-Based Models

- ABM with N agents and L sites is a large Markov chain $(X_t, t \in \mathbb{Z})$, where $X_t = (X_{t,1}, \dots, X_{t,N})$ and with alphabet size L^N
- Quantity of interest at t is the number of agents at site l, i.e. $Y_t = (Y_{t,1}, \dots, Y_{t,L})$, where

$$Y_{t,l} = f_l(X_t) = \sum_{n=1}^{N} \mathbb{1}_{\{X_{t,n}=l\}}$$

- $(Y_t, t \in \mathbb{Z})$ is a function of a Markov chain:
  - Has alphabet size smaller than $N^L \ll L^N$ if $N \gg L$
  - Not Markov in general, but under certain conditions on $(X_t, t \in \mathbb{Z})$ and $f_1, \dots, f_l$ → *lumpability*[6]
- Conjecture: $(Y_t, t \in \mathbb{Z})$ is Markov if agents do not interact
- Consequence: "Small" Markov chain $(Y_t, t \in \mathbb{Z})$ *probabilistically equivalent* to ABM
- Next Steps/Issues:
  - Verify conjecture and generalize to lumpability of ABMs with interacting agents[7]
  - Transform agent rule set to Markov transition probabilities, e.g.
    - Agents leave site l Bernoulli-distributed, $Bern(p)$
    - Number of agents at site l at time $t+1$ are binomially distributed, $B(Y_{t,l}, 1-p)$
    - Potential problems with large binomial coefficients → approximations

Community Detection in Graphs

- Social networks represented as a graph $G = (V, E)$ with a set of nodes V and a set of edges E
- Node set is large → *Parallelise community detection*:
  - For the neighbourhood $G_n = (V_n, E_n)$ of each node n, find the community structure $C_n = \{c_{n,1}, \dots, c_{n,j}\}$ that maximises modularity (Louvain algorithm[8]):

$$\sum_{c \in C_n} \left[\frac{|E(c)|}{|E_n|} - \left(\frac{\sum_{v \in c} \deg(v)}{2|E_n|} \right)^2 \right],$$

    where $E(c)$ is the subset of E for which all endpoints lie in c
  - Eliminate all communities $c_{m,k}$ that are subsets of other communities
  - Merge pairs of overlapping communities $c_{m,k}$ and $c_{l,j}$ if

$$\lambda_{G(c_{m,k} \cup c_{l,j})} > max\{\lambda_{G(c_{l,j})}, \lambda_{G(c_{m,k})}\},$$

    where $\lambda_{G(c)}$ is the second smallest Laplacian eigenvalue of the graph $G(c)$

References

[1] https://www.who.int/airpollution/en/
[2] Horváth Z., Liszkai B., Istens G., Zsebők P., Szintai B., Rácz E., Környei L., Harmati I. (2016) "Integrated urban air pollution dispersion modelling framework and application in air quality prediction of the city of Gyor", Proc. 17th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Budapest.
[3] Suleimenova D., Bell, D. & Groen D. (2017) "A generalized simulation development approach for predicting refugee destinations", *Scientific Reports*, 7:13377.
[4] Suleimenova D., Bell D. & Groen D. (2017) "Towards an automated framework for agent-based simulation of refugee movements", in *Proc. IEEE Winter Simulation Conf.* pp. 1240–1251. Las Vegas, Nevada.
[5] Hess M., Alla A., Quaini A., Rozza G., Gunzburger M. (2018) "A localized reduced-order modeling approach for PDEs with bifurcating solutions", in *Computer Methods in Applied Mechanics and Engineering*, 351:379-403.
[6] Kemeny J. G. & Snell J. L. (1976) Finite Markov Chains, Springer.
[7] Banisch S. (2016) Markov Chain Aggregation for Agent-Based Models, ser. Understanding Complex Systems, Springer.
[8] Blondel V. D., Guillaume, J-L., Lambiotte, R., Lefebvre, E. (2008) "Fast unfolding of communities in large networks", in *Journal of Statistical Mechanics: Theory and Experiment*, 10:P10008.

Project #824115 funded by the Horizon 2020 Framework Programme of the European Union.