Coalition Formation Algorithm of Prosumers in a Smart Grid Environment

Nicolas Gensollen, Vincent Gauthier, Michel Marot and Monique Becker
CNRS UMR 5157 SAMOVAR,
Telecom SudParis/Institut Mines Telecom
Email: {nicolas.gensollen, vincent.gauthier, michel.marot, monique.becker}@telecom-sudparis.eu

Abstract—In a smart grid environment, we study coalition formation of prosumers that aim at entering the energy market. It is paramount for the grid operation that the energy producers are able to sustain the grid demand in terms of stability and minimum production requirement. We design an algorithm that seeks to form coalitions that will meet both of these requirements: a minimum energy level for the coalitions and a steady production level which leads to finding uncorrelated sources of energy to form a coalition. We propose an algorithm that uses graph tools such as correlation graphs or clique percolation to form coalitions that meet such complex constraints. We validate the algorithm against a random procedure and show that, it not only performs better in term of social welfare for the power grid, but also that it is more robust against unforeseen production variations due to changing weather conditions for instance.

I. INTRODUCTION

One of the key ideas in the smart grid revolves around the introduction of communication means inside the power grids. This could enable complex improvements in the energy management and leads progressively to a greener energy system [1] [2]. Distributed energy resources (DER) such as wind turbines or photovoltaic panels will probably not only appear in remote farms, but also in residential areas. Together with electric vehicles, and demand side management programs [3], they will constitute the building blocks which will help to turn the today pure energy consumers into true actors of the grid operation [1]. Such agents that both consume and produce energy are ready to make concessions (appliances delays, V2G) to ensure grid stability, and are commonly called prosumers [1], [4].

There seems to be a clear consensus on the benefit of having bidirectional communication flows between the prosumers and the grid, as demand side management relies on such an architecture [2]. Nevertheless as the number of active prosumers is expected to rise, it is safe to assume there is a need for more complex communication patterns (prosumers grouping together into so-called "coalitions") that will help to decrease the communication burden and satisfy the multiple requirements of the power grid management.

Formation of coalitions inside a smart grid environment could be applied to various types of agents such as self sufficient micro-grids [5], sizable and adjustable virtual power plants (VPP) [1], [6], or fleet of electric vehicles that back up the grid in emergency situations (V2G) [1]. These are only a few use cases where the coalition of multiple agents can enhance the grid reliability and in the mean time reduce the communication burden.

In this paper, we consider the case where prosumers coalesce into virtual groups (the so-called coalitions) in order to be able to produce enough power which may be sold to the grid. If the coalition's production is sufficiently high and stable, the coalition can negotiate on a day by day basis with the grid by announcing each day an expected production for the next day.

We focus on how to statistically improve the production stability by carefully forming coalitions of prosumers that have greater probabilities of staying in acceptable range of production. By using prediction techniques, we form coalitions that meet the contract values ranges proposed by the grid, enabling it to schedule its production accordingly. Meeting contract values for the prosumers is especially relevant in power trade conditions, where energy is traded based on day-ahead predictions.

In this context, participants should try to minimize their prediction errors in order to maintain a stable state for the grid. They may even occur some penalties if the productions deviate from the initial contract values. However, it remains pertinent for both side to maintain the production as stable as possible : for the prosumer, the stability (with renewable energy sources) will impact its net gain, and for the power grid, it will help to maintain its reliability.

In this paper, we seek to form coalitions that are able to announce "high contract values with high reliability". We thus aim at:

- Building a realistic prosumer model with renewable production sources based on weather data (see section III-A).
- Define a coalition formation model that enables the grid to set some requirements under which any group of prosumers will be allowed to sell its production (see section III-B)
- Define a utility function that will satisfy the grid requirements and maximize the stability of the prosumer coalition productions (see section III-C)
- Define a coalition formation algorithm (see section IV)

In section III, we consider that the agent's production depends on meteorological conditions (wind turbines and photovoltaic panels (PV) as generators), various energy mix and preferences, and their appliances (loads). We use meteorological data (see section III-A) to account for seasonal

and daily variations in the prosumer's energy profiles as well as realistic geographical correlations between different agents. With these ingredients, we are able to run realistic simulations and record the different output profiles as time-series.

In order to select groups of prosumers that lower the volatility of the coalition's productions, we define an algorithm based on a popular approach for stocks market clustering [7]. We use a modified clique percolation algorithm (see section IV) that will enable us to expand the cliques as needed in order to form proper coalitions that fulfill both the grid requirements and lower the production volatility. Finally, section V provides some results and a conclusion in section VI.

II. RELATED WORK

Traditionally, forming coalitions in a pool of agents can be done either in a centralized way where a single central unit is responsible for all the computations or in a distributed way where agents have only local knowledges and take actions accordingly. It is of common use to represent the situation and assess the stability of the solution by using game theoretic tools. Some papers [8] [9] focus then on finding an optimal coalition structure giving a pool of autonomous self interested agents using distributed merge/split algorithms.

Tackling the stability issues of renewable DER, the TradeWind project [10] simulated the impact of wind power on electricity exchange and cross-border congestions by using a flow-based market model. The idea revolves around identifying key european interconnections (already existant or not) in order "to make optimal use of the european spacial de-correlation of wind power". It was indeed shown that geographical aggregation provides smoothening effects and that the amount of prediction errors for wind power in a geographical region diminishes as the region size increases, especially for short forecast horizons.

On a narrower scale, [11] study the formation, in a game theoretic setting, of virtual power plants (VPP) composed of multiple self-interested DER. Two requirements for the formation of virtual power plants are considered : the reliability of supply and the minimization of entities the grid has to deal with. From this, [11] builds a pricing mechanism that encourages VPP to report true estimates of their aggregated production and penalizes prediction errors. A redistribution scheme of the VPP to the DER is also constructed such that the payoff allocation lies in the core of the game, meaning that no DER has an incentive to leave the coalition.

In this paper we do not consider stability against player defection but focus on a statistics-oriented definition of coalition stability. We want to form coalitions based on utilities that depend on statistical properties of time-series values. The setting is thus similar with some financial studies on stock exchanges, where researchers tried to find relevant correlated clusters of stocks based on their daily prices variations. There exists well-known algorithms such as K-means or hierarchical clustering that are traditionally used for such purposes. Nevertheless, despite their popularity, they do not appear particularly well

suited for meaningfully translating complex correlation relationships between time series while clustering them, especially when these series exhibit complicated underlying patterns. We will show later in this paper that using K-means as a clustering technique for grouping prosumers together leads to less good results (see fig. 5).

In [7] the author introduced an approach where stocks time-series of their daily log returns, are organized in a graph such that stocks exhibiting similar price fluctuation patterns are close to each other. This closeness notion is formalized with a similarity measure based on Pearson correlation coefficient (see notations in table I) ($d_{ij} = \frac{1}{2}\sqrt{2(1-\rho_{ij})}$ or $d_{ij} = 1-\rho_{ij}^2$) that enables to weight the edges of the graph.

ϵ-graphs, consists in filtering edges based on their weight, only keeping edges whose weights are less than ϵ. In [12], [13], the authors studied the topological properties (average clustering, connectivity, relative number of cliques) of the correlation graph against those of growing random graphs, depending on the threshold ϵ. However, there is no well defined method to select a right tradeoff ϵ as a function of the network topology.

Presented in this way, the time-series clustering task seems very close to graph community detection. Communities in networks are indeed often seen as groups of nodes exhibiting high internal densities of links as well as a low density across communities, and several topology oriented techniques for finding communities are present in the litterature ([14] [15] [16]).

For our purpose where decorrelation is the closeness notion, such algorithms tend to have some difficulties because only a few inclusions of very high correlations can strongly affect the stability of a coalition. Furthermore, these graphs tend to exhibit far less clustering than their correlation-based counterparts, which can increase the difficulty of finding communities.

However, we will see in the following, that a clique (a complete subgraph) in our filtered decoralation-based graph exhibit, by construction, low correlation. We indeed removed in the filtering (see ϵ-graphs) all links that stands for high correlation. There exists greedy heuristics based on cliques in the litterature such as clique percolation [17] which realizes local optimizations of a fitness function and results in overlapping community structures.

Detection of overlapping communities is actually a very active field of research, especially in social networks where a person might belong to several communities. Section IV will explain how we used such techniques on the decorelation graph in order to form coalitions of prosumers.

III. MODEL

A. Prosumer model

Our major concern is to characterize a prosumer behavior with realistic patterns of consumption, production as well as realistic geographical correlations between them. Only wind turbines and PV (photovoltaic panels) will be used as generators throughout this paper, and we denote by $\mathcal{A} = \{a_1, ... a_N\}$

the set of prosumers. To account for the production of a prosumer, we rely on French's weather data from 2006 to 2012 sampled every 3 hours [18] (similar data are available for the United States covering 2010 [19]).

As shown in the first part of figure 1, the first step consists in discretizing the studied zone around well chosen weather stations and gathering three kinds of traces (wind speed, cloudiness, and temperature) that we organize in a structure we will refer to as "climate vector" in the following. We consider that climate vectors are constant over their area, meaning that, if two agents are in the same area, they are exposed to the same climate vector. We use existing DER models enabling conversion between weather variables and output powers (see power curves in [11] [20]). [21] provides also a convenient way of using cloudiness traces as realistic degradation factors in a *"clear blue sky"* solar model, which enables us to rebuild realistic solar irradiance traces for stations that do not provide such information directly.

Consumption behaviors are characterized by a model we designed, which takes into account two major cycles :

- **Daily cycles** : Consumption is low during night, and higher during the day with two picks in the morning and evening. Some noise is added so that prosumers have similar but not identical cycles.
- **Seasonal cycles** : Consumption is higher in the winter because of heating and low in the summer (air conditioning is not considered). Temperature traces are used for modeling these cycles.

At this point, modeling agents consists in fixing a few parameters (most of the time drawn from random distributions) such as the geographical position, the number of wind turbines, PV, appliances, the temperature of comfort, and so on. The objective is that, depending on the weather of his zone, the DER and appliances he owns, and the way he decides to heat his home, a given prosumer is able to compute his production/consumption at any time.

Let $P_i(t)$ be the instantaneous available power of agent i at instant t (his instantaneous production minus his instantaneous consumption). During the simulation (from t_0 to t_K), all agents record their values with an hour time interval. As expected, the prosumer's time-series exhibit high seasonal patterns, but completely different from agent to agent because they depend both on the energy mix and the habits. Nevertheless, these macro variations can hide the interesting information in the correlation coefficients and twist the rest of our process. Thereby, we remove these seasonal effects for all agents (see second block of figure 1). We denote by $\mathcal{T}_i = \{P_i(t_0), ..., P_i(t_K)\}$ the resulting time-series for agent i, and by $\mathcal{F}(P_i)$ the corresponding probability distribution.

B. Coalition formation model

Let us recall that our goal is to group prosumers into coalitions so that the global power production resulting of the superposition of individual prosumer's productions be
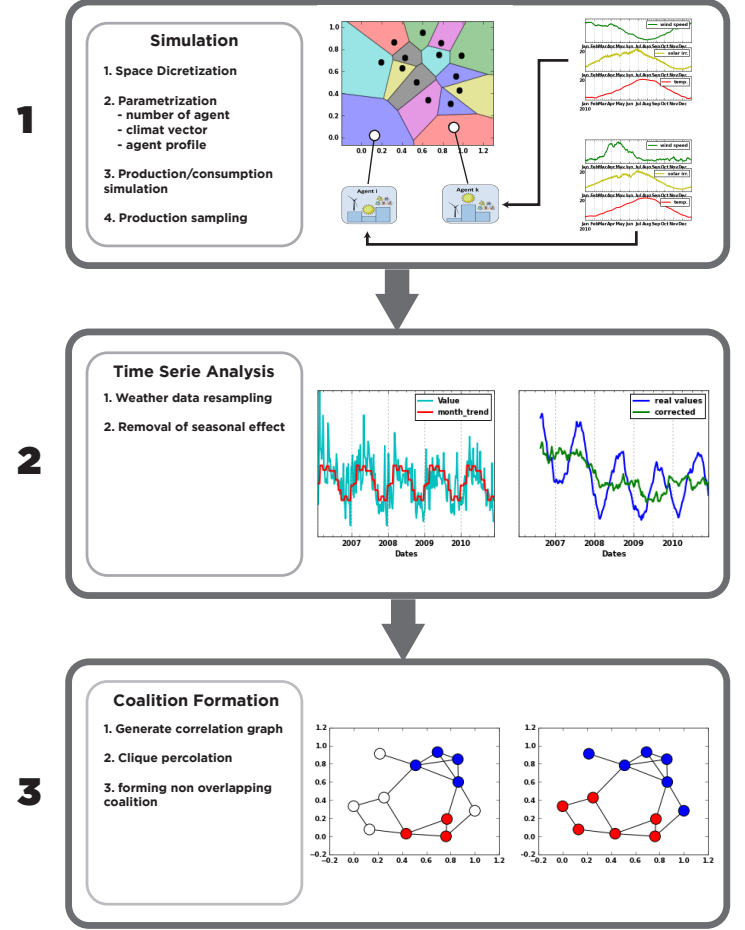

Figure 1: Process diagram

sufficiently predictible and above a minimum power level. We thus extend the agent notations to any coalition $s \subset \mathcal{A}$:

- $P_s(t) = \sum_{i \in s} P_i(t)$
- $\mathcal{T}_s = \{P_s(t_0), ..., P_s(t_K)\}$ with $\mathcal{F}(P_s)$ the corresponding probability distribution

Furthermore, we consider that a coalition *s* has the possibility to announce a contract value P_s^{CRCT} on the market. Due to the volatile nature of DER productions or user consumptions, the available power of a coalition will oscillate around P_s^{CRCT}. It is of the responsibility of the information system that forms coalitions to select agents in order to provide a production as stable as possible. The probability of *s* value being below its contract value $Pr[P_s \leq P_s^{CRCT}]$ appears therefore as a quantity of interest that we would want to keep as low as possible.

To illustrate what is done in the following, let us consider a very simple toy example with two agents, say *i* and *j*, with production values following gaussian probability distributions

$\mathcal{F}(P_i) = \mathcal{N}(\mu_i, \sigma_i)$ and $\mathcal{F}(P_j) = \mathcal{N}(\mu_j, \sigma_j)$, such that the joint probability distribution $\mathcal{F}(P_{ij})$ of the coalition $\{ij\}$ is also a gaussian with the following parameters :

$$\begin{cases} \mu_{ij} = \mu_i + \mu_j \\ \sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2 + \rho_{ij}\sigma_i\sigma_j} \end{cases} \tag{1}$$

ρ_{ij} is the Pearson correlation coefficient between P_i and P_j (see table I for notations). Because all distributions are gaussians, we can easily write the probability $Pr[P_{ij} \leq P_{ij}^{CRCT}]$ as :

$$\frac{1}{2}\left[1 + erf\left(\frac{P_{ij}^{CRCT} - \mu_{ij}}{\sigma_{ij}\sqrt{2}} \right) \right] \tag{2}$$

Let $\phi \in [0,1]$ be the **reliability** constraint imposed to the coalition $\{ij\}$. It stipulates that if $\{ij\}$ wants to join the market, the probability of $\{ij\}$ value (at any instant t) being below its contract value should at most be ϕ. That is, $Pr[P_{ij} \leq P_{ij}^{CRCT}] \leq \phi$. We also impose to $\{ij\}$ a **minimum contract value** constraint denoted by P^{MIN}, which states that the coalition should announce more than P^{MIN} to be accepted ($P_s^{CRCT} \geq P^{MIN}$).

Coalition $\{ij\}$ has thus potentially a set of possible contract values that respects both rules. The strategy of $\{ij\}$ (and all other coalitions in the paper) consists then in choosing, if it exists, the maximum value of this constrained set. We denote by $P_\phi(s)$ this maximum value for a coalition s. For coalition $\{ij\}$ in the toy example, $P_\phi(s)$ can be computed using eq. 2 : $P_\phi(ij) = \mu_{ij} - \sigma_{ij}\sqrt{2}erf^{-1}(1 - 2\phi)$. If $P_\phi(ij) \geq P^{MIN}$, then $\{ij\}$ can announce $P_{ij}^{CRCT} = P_\phi(ij)$ on the market, else it is not allowed to enter. It thus appears (as it was intuitively understandable) that, for equivalent sizes, coalitions with low relative standard deviations (σ_{ij}/μ_{ij}) are able to announce higher contract values.

In the following, we do not make any gaussian assumption as in the previous toy example (see for instance the densities at the coalition level on figure 4). We investigate then the developement of a heuristic that tries to find a compromise between maximizing the production sold on the market (benefits) on one hand (see eq. 3) and minimizing the probability that the coalitions fail to sustain their contract values (stability) on the other hand (see eq. 4). Table I summarizes the main notations.

$$\underset{\substack{S \subset CS \\ |S| = N_{COAL} \\ \forall s \in S, \, |s| \neq 0 \\ P_s^{CRCT} \geq P^{MIN}}}{\arg\max} \sum_{s \in S} P_s^{CRCT} \tag{3}$$

$$\underset{\substack{S \subset CS \\ |S| = N_{COAL} \\ \forall s \in S, \, |s| \neq 0 \\ P_s^{CRCT} \geq P^{MIN}}}{\arg\min} \sum_{s \in S} \Pr\left[P_s \leq P_s^{CRCT} \right] \tag{4}$$

This last minimization objective is explored by organizing prosumers such that the synergy terms of the standard deviations ($\rho_{ij}\sigma_i\sigma_j$ in the example above) are minimized. As

d_{ij}	Distance metric between timeseries i and j
ρ_{ij}	Pearson's correlation coefficient between timeseries i and j
$\mathcal{A}$	Agent set
$P_i(t) \in \mathbb{R}$	Instantaneous production of agent i at time t
$P_s(t)$	Instantaneous production of coalition s
$\mathcal{F}(P_i)$	Probability distribution of i's production
$\mathcal{F}(P_S)$	Probability distribution of coalition s production
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ
$\phi \in [0,1]$	Reliability threshold
$P^{MIN} \in \mathbb{R}^+$	Min value threshold
$P_s^{CRCT} \in \mathbb{R}^+$	Contract value announced by coalition s
$P_\phi(s) \in \mathbb{R}^+$	Max contract value under ϕ constraint
$\mathcal{U}_{\phi,\ PMIN}(s)$	Utility function
$\tau_i(s) \in \mathbb{R}$	Overlapping to disjoint coalitions mapping
$\epsilon \in [0,1]$	Pruning threshold for the correlation graph
$G_\epsilon(\mathcal{A}, E)$	Filtered correlation graph
$N_{COAL} \in \mathbb{Z}^+$	Number of desired coalition
CS	Set of all coalition structures
$\Theta(G)$	Set of non overlapping cliques in graph G
$\mathbf{1}_X$	Indicator function over set X

Table I: Notations

explained in the example above, if a coalition s wishes to join the market, it has to choose a contract value within a constrained set. Only coalitions whose set is not empty are able to enter the market (valid coalitions). For simplification, we consider that valid coalitions will always apply the same economically consistent strategy of announcing the highest possible contract value of this set ($P_s^{CRCT} = P_\phi(s)$). Let us define a valid coalition more formally as a coalition satisfying the following rules :

$$\begin{cases} Pr[P_s \leq P_\phi(s)] \leq \phi \text{ \textit{(reliability rule)}} \\ and \ P_\phi(s) \geq P^{MIN} \text{ \textit{(min value rule)}} \end{cases} \tag{5}$$

C. Utility function

Adressing the problem of eq.3 and 4, we choose a utility function (eq. 6) that derives directly from the above remarks. If a coalition cannot provide a valid contract value, it receives naturally a utility of zero. Furthermore, the utility increases with the overall production (sum of the contracted values) as stated in the objective of eq.3. The stability objective described in eq.4 is also present in this utility function through the ϕ parameter. Finally, the $1/|s|$ term in eq. 6 indicates that we favor small coalitions, mainly because they are easier to maintain.

$$\mathcal{U}_{\phi,\ PMIN}(s) = \mathbf{1}_{s\ valid}\frac{P_\phi(s)}{|s|} \tag{6}$$

Obviously, maximizing this utility function amounts to maximizing the coalition contract value with the minimum possible number of agents. In such settings, and through the clique percolation procedure (section IV), we will show that coalitions with high utilities can be computed.

IV. COALITION FORMATION

This section explains the process with which we form the coalitions (see the third block of Figure 1). First, we need to simulate the time-series of available power (first two blocks of

Figure 1). We consider a pool $\mathcal{A}$ of 200 agents, whose parameters were chosen randomly. The prosumers are positioned (also randomly) on a square lattice previously filled with climate vectors obtained from the french data sets (see section III). Simulations were run from february 2006 to december 2010 such that we are dealing with 200 hourly sampled time-series of available power over this period. Removing season trends finally leads us to the formation of coalitions.

The model we used in order to simulate time-series of available production provides some diversity because of the combination between the energy mix (renewable generators combination) and climate vectors. Nevertheless, as the number of agents grows, the time-series tend to exhibit similar patterns. This is apparent when creating a correlation graph with similar metric such as defined in [12] or [13] $(1 - \rho_{ij}^2)$, where well defined clusters appear in the ϵ-graph for any values of ϵ.

However, these clusters of strongly correlated time-series are the exact opposite of what we are seeking. We can indeed consider them directly as coalitions and compute their utilities, and the results show (see the green curves with diamonds on figure 5), as expected, terrible values. We will indeed consider in section V this case as a worst case scenario, far worse than a trivial random split of the agents in the same number of coalitions (see the red curves with triangles of the random case on figure 5).

We thus opt for reversing the metric (ρ_{ij}^2) such that uncorrelated time-series are close to each other in the graph and correlated time-series are distant. We defined a graph $G_\epsilon(\mathcal{A}, E)$ where $\mathcal{A}$ is the set of agents and E the edge set. A given link between i and j is present if $\rho_{ij}^2 \leq \epsilon$, that is, $e_{i,j} = \mathbf{1}_{\{\rho_{ij}^2 \leq \epsilon\}} \rho_{ij}^2$. As expected [13], independently of the ϵ parameter, G_ϵ exhibits henceforth much less clustering and communities seem hardly visible. Therefore, using classical clustering or community detection algorithms seem to provide poor results.

However, as seems intuitively understandable, cliques of this graph tend to exhibit very good utility values. Such structures contain indeed a link between every two nodes, meaning that the overall correlation is quite small. Obviously, the ϵ parameter is indirectly responsible for the sizes of the cliques : if it is too low, G_ϵ does not provide enough cliques, conversely, if ϵ is too high, we loose important information as the graph becomes very dense and cliques tend to overlap strongly. For large values of ϵ and for non trivial number of agents, finding cliques can even become computationally intensive.

Despite being direct and simple, improving the sizes of the cliques by increasing ϵ does not seem elegant. Furthermore, as the utility function is also focused on maximizing $P_\phi(s)$, it might be the case that some cliques benefit from additional agents even if they don't form a clique anymore. Cliques for small values of ϵ appear thus as good seeds for forming stable coalitions rather than directly acceptable coalitions. We define N_{COAL} as the desired number of coalitions fixed by the user. Furthermore, let $\Theta(G_\epsilon)$ be the set of non overlapping cliques in G_ϵ, we choose the ϵ parameter as the smallest value such that G_ϵ contains at least N_{COAL} non overlaping cliques :

$$\epsilon^\star = \min_{\epsilon \in [0,1]} \left\{ \epsilon \ s.t \ |\Theta(G_\epsilon)| \geq N_{COAL} \right\} \quad (7)$$

We construct $G_{\epsilon^\star}$ such that it contains at least N_{COAL} potential seeds that may or may not be above the grid requirements, but exhibit low correlations among the members. Naturally, grid requirements could eventually be reached and social welfare increased by incorporating more nodes into the seeds. For this purpose, we use a clique percolation algorithm on $G_{\epsilon^\star}$ that will extend the seeds over a correlation constrained environment. The main idea behind this algorithm is to make the seeds grow by local optimization of a fitness function, usually based on internal and external degrees of the seeds [17]. The main difference between our case study and this algorithm for community detection is that we chose the fitness function as our utility function (see eq.(6)) that both maximizes the production and minimizes its volatility (see eq.3 and eq.4).

As explained in section II, clique percolation leads generally to overlapping communities. For simplicity, in this paper, we wish to keep the coalitions disjoint and leave the management of overlapping coalitions for future work. We thus implemented a simple heuristic that considers nodes in multiple seeds one by one and chooses its final coalition as the one that "needs it the most" in terms of utility loss. More formally, for a coalition s and a node $i \in s$, we define[1] :

$$\tau_i(s) = \frac{\mathcal{U}_{\phi, \ P^{MIN}}(s) - \mathcal{U}_{\phi, \ P^{MIN}}(s - \{i\})}{\mathcal{U}_{\phi, \ P^{MIN}}(s)} \quad (8)$$

If node i belongs to multiple coalitions, the only coalition including node i is the one that maximizes τ_i. At this point, we have three degrees of freedom : the reliability (ϕ), the required power to enter the market (P^{MIN}), and the number of desired coalitions (N_{COAL}). Next section shows the utility behavior within the parameters space $\{\phi, P^{MIN}, N_{COAL}\}$ and provides results of our 200 agents test.

V. RESULTS

As shown in figure 2a, the parameters ϕ and P^{MIN} shape the utility function such that, if ϕ is close to zero, the reliability requirement is very high. Only small values of P^{MIN} could then potentially lead to valid coalitions (and positive utilities). Conversely, the higher ϕ, the less constraints are imposed to the coalitions and more valid coalitions could be formed for a larger spectrum of P^{MIN}. Obviously, the highest utility values are found for high ϕ, because the formed coalitions are able to announce higher contract values, yielding higher utilities.

In figure 2b, we fix the reliability to a given empirical value ($\phi = 0.08$). We can observe how the social welfare $\sum_{s \in S} \mathcal{U}_{\phi, \ P^{MIN}}(s)$ evolves according to P^{MIN} and N_{COAL}. In our setting we notice a maximum point around $N_{COAL} = 3$, followed by a decrease. There is clearly a maximum number of coalitions sustainable for a given set of parameters. After that value the coalitions newly formed won't be able to pass the

---

[1]Not to be mingled with the Shapley value concept

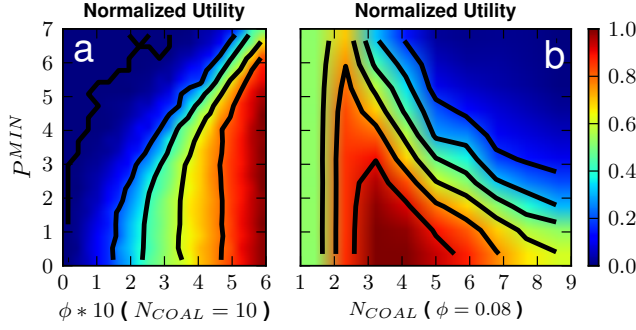Figure 2: Utility as function of the parameters space $\{\phi \in [0,1], N_{COAL} \in \mathbb{Z}^+, P^{MIN} \in \mathbb{R}^+\}$: a) In $\{\phi, P^{MIN}\}$ when N_{COAL} is fixed. b) in $\{N_{COAL}, P^{MIN}\}$ for a fixed ϕ. For readability, P^{MIN} is expressed in tenth of MW.

grid requirements. Moreover, reckon that increasing N_{COAL} means also increasing ϵ, leading to denser graphs, meaning that the algorithm performance will also decrease.
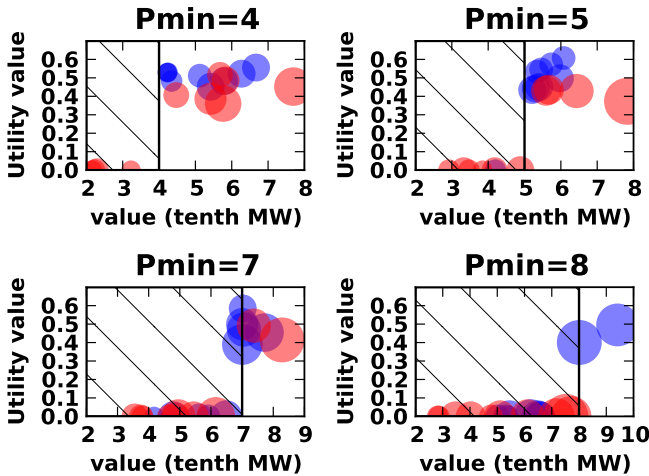

Figure 3: Evolution of the coalitions for different values of P^{MIN} for clique percolation (blue dots) and random process (red dots). The diameters of the dots are proportional to the coalition sizes. The hatched zero utility zone corresponds to the *"under-requirement space"*.

In fig. 3, we perform a simulation with 200 prosumers, and we fix the reliability $\phi = 0.1$ meaning that coalitions should produce more than their contract values at least 90% of the time. We observe the result of our algorithm for a fixed number of coalitions of $N_{COAL} = 10$ and let the algorithm form the coalitions as the grid changes according to P^{MIN}. We compare our algorithm with a random process that splits the agents in N_{COAL} coalitions. Fig. 3 shows this evolution for our algorithm (blue dots) and for the random process (red dots). The diameter of a dot is proportional to the number of agents in the coalition and the higher the dot, the higher its utility. The P^{MIN} values of the x axis are expressed in
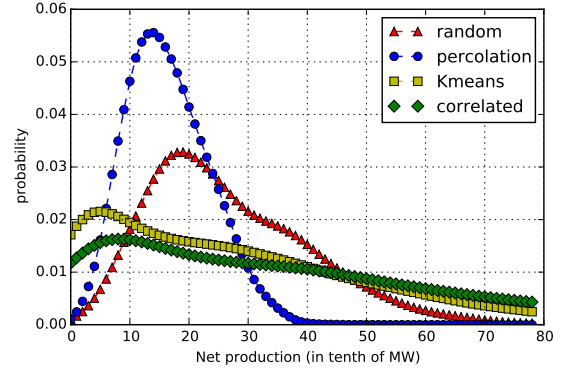

Figure 4: Production probability densities for four coalitions. One obtained with our algorithm based on clique percolation (blue curve with circles), one with a random formation (red curve with triangles), another one with a K-means approach (yellow curve with squares), and finally one with correlated agents (green curves with diamonds).

tenth of MW for readability and the hatched zone corresponds to the *"under-requirement space"*, meaning that whenever a coalition is in this zone, it has a null utility. Looking at fig. 3, we can see first that our algorithm performs better than the random formation algorithm at finding high valued coalitions. The blue dots allowed to enter the market (non hatched zone) are indeed outnumbering the red dots, especially when the grid requirements are neither too low nor too high ($P^{MIN} = 5$ in fig. 3).

In the following, we consider and compare four different methods : our method based on clique percolation (blue curve with circles), the previous random aggregation of agents (red curve with triangles), a K-means clustering approach (yellow curve with squares), and correlated coalitions (green curves with diamonds) obtained with a correlation graph and that stands for a "*worst case scenario*".

As visible on figure 4, the production densities at the coalition level exhibit very different shapes. While our algorithm results in a nice density with a low (compared to the other curves) standard deviation, the other methods seem to form coalitions that will have a very large spectrum for their production. In more details, figure 5a presents the evolution of social welfare as the number of coalitions increases (all other parameters are kept constant) for the same four methods. As for figure 5b, it shows the percentage of coalitions able to enter the market for different values of P^{MIN}. For consistency, we average the results of both plots over 100 realizations (colored zones around the curves shows the standard deviations). When the grid requirements are constant (fig 5b), and the number of desired coalitions is low, clique percolation performs better than random case. Moreover, when N_{COAL} gets bigger, the performance of a random split tumbles down rapidly while our clique percolation is able to maintain efficiently the social welfare of the coalitions formed.
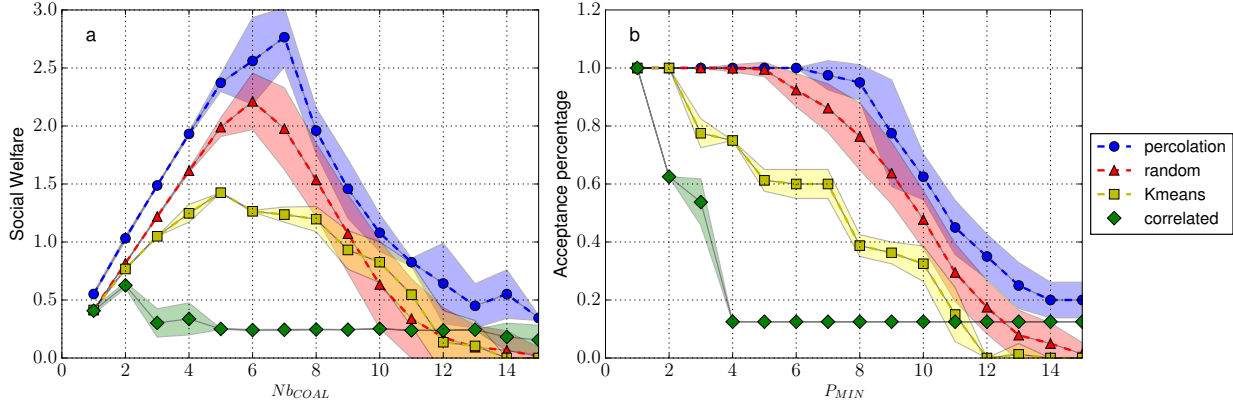
Figure 5: Evolution of the social welfare when the number of coalitions is varying **a)**, and the percentage of coalitions able to enter the market for different values of P_{MIN} **b)**. Blue curves with circles and red curves with triangles represent respectively the results for clique percolation and the random process. For comparison, the green curves with diamonds and the yellow ones with squares show respectively the results obtained with coalitions of correlated agents and with a K-means clustering.

When the grid requirements vary, for very low P^{MIN}, all coalitions for all algorithms are able to enter the market, yielding an acceptance percentage of 100%. But as P^{MIN} increases, we see the percentage of the correlated coalitions going down very quickly.

After a few increases in P^{MIN}, the percentage of the random procedure starts dropping while it stays almost constant for our algorithm. For $P^{MIN} = 8$, we see that 100% of the coalitions are able to enter with our algorithm while approximately 75% of them enters for the random case (only 40% and 15% for the other cases). Finally, when the grid requirements become too high, the acceptance percentage of our algorithm tends slowly to zero (not shown in this plot for readability).

VI. CONCLUSION

We believe that the originality of this paper lies in its willingness to exploit the de-correlation of prosumer profiles in order to build stable coalitions. In this direction, we presented a model based on meteorological traces, that captures the complex "*energy mix/climate vectors*" combination and generates realistic production and consumption patterns. We then built a framework that enables the grid to specify stability and minimum production requirements for filtering the coalitions. On this basis, we proposed a simple algorithm that seeks for uncorrelated prosumer patterns as potential seeds and expand them in coalitions able to rise above the grid requirements. We validated our algorithm against a random choice of coalitions, which can be considered as an average case. We showed that it performs better (coalitions are more stable and the overall production is more important) and that it exhibits a higher robustness/flexibility against grid requirements changes. The worst case scenario is represented by coalitions of correlated prosumers.

Interesting leads for future works would be the use of correlated clusters to reduce the number of entities the algorithm has to deal with, or the introduction of a payoff allocation towards the prosumers such that the stability against player defection could be analyzed through game theory. Showing that maintaining the coalitions formed with our algorithm necessitates less communication and less storage capacity could also lead to a stimulating project. Besides, we believe

that not restricting the algorithm to non overlapping coalitions and studying the strategies and weights of nodes with multiple options could lead to interesting research.

REFERENCES

[1] S. D. Ramchurn *et al.*, "Putting the 'smarts' into the smart grid: A grand challenge for artificial intelligence," *Commun. ACM*, vol. 55, no. 4, pp. 86–97, Apr. 2012.

[2] C. Wu *et al.*, *Smart Grid Communications and Networking*, E. Hossain *et al.*, Eds. Cambridge University Press, 2012.

[3] P. Samadi *et al.*, "Utilizing renewable energy resources by adopting DSM techniques and storage facilities," *International Conference on Communications*, pp. 4221–4226, 2014.

[4] L. Xiao *et al.*, "Anti-cheating prosumer energy exchange based on indirect reciprocity," in *International Conference on Communications*, 2014, pp. 599–604.

[5] S. Pahwa *et al.*, "Topological Analysis and Mitigation Strategies for Cascading Failures in Power Grid Networks," 2010.

[6] M. Braun, "Virtual Power Plants in Real Applications Pilot Demonstrations in Spain and England as part of the European project FENIX," 2009.

[7] R. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.

[8] W. Saad *et al.*, "A distributed coalition formation framework for fair user cooperation in wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 9, pp. 4580–4593, September 2009.

[9] X. Luan *et al.*, "Cooperative power consumption in the smart grid based on coalition formation game," in *Advanced Communication Technology (ICACT)*, Feb 2014, pp. 640–644.

[10] F. Van Hulle *et al.*, "Integrating Wind, European project report," 2009.

[11] G. Chalkiadakis *et al.*, "Cooperatives of distributed energy resources for efficient virtual power plants," in *The 10th International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 787–794.

[12] A. Garas *et al.*, "The structural role of weak and strong links in a financial market network," *The European Physical Journal B*, vol. 63, no. 2, pp. 265–271, 2008.

[13] J.-P. Onnela *et al.*, "Clustering and information in correlation based financial networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 353–362, 2004.

[14] M. E. J. Newman, "Spectral methods for network community detection and graph partitioning," p. 11, Jul. 2013.

[15] M. Girvan *et al.*, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[16] M. E. J. Newman, "Community detection and graph partitioning," no. 2, p. 5, May 2013.

[17] A. Lancichinetti *et al.*, "Detecting the overlapping and hierarchical community structure in complex networks," *New Journal of Physics*, vol. 11, no. 3, p. 033015, 2009.

[18] [Online]. Available: http://www.infoclimat.fr

[19] [Online]. Available: http://www.ncdc.noaa.gov/

[20] [Online]. Available: http://www.wind-power-program.com

[21] G. Dans *et al.*, "Rayonnement solaire," vol. 360, 2007.