The "Digital Territory" as a complex system of interacting agents, emergent properties and technologies

Yannis C. Stamatiou

Based on work jointly done with V. Liagkou and E. Makri and P. Spirakis

September 7, 2007

The Digital Territory (DT)

- Artificial Intelligence: The focus is towards emergence of "intelligence" in a single, complex individual

- Artificial Life: The focus is the emergence of global "intelligence" based on interactions of simple individuals

- A new "artificial" concept, the *Digital Territory*: this concept integrates Artificial Life with Artificial Intelligence – it describes worlds with moving agents which, however, move in complex terrains which contain elements of both the physical and digital world as well as "real" intelligence since it integrates devices with human beings in a complex pattern of interactions.

Defining a DT

Definition. *(an attempt ...) A DT is a spherical region in some d-dimensional space (e.g. 2-dimensional or 3-dimensional Euclidean spheres) composed of moving entities which appear or disappear in an unpredictable manner (uniformly at random) within the region. Around each of the entities another region is defined composed of the space points lying within a prespecified distance from the entity (this models the sensing and communication capabilities of the entity).*

We are interested in describing the activities of the entities within this region using a mathematical formalism and mapping the elements of the formalism into currently available technological devices.

The technological basis of DTs

- Mobile devices
- 4G (harmonizing 3G, WiFi, WiMax, and Bluetooth)
- Location awareness
- Sensor technology (Radiation, Gas, Temperature, Mechanical strain, Position and Proximity, Sound, Magnetic field, Humidity, Speed, Acceleration etc.)
- RFID (Radio Frequency IDentification)
- Wearable computing devices
- Bio-implants

Modeling Appearance, Proximity and Interactions - Random Graphs models

In what follows, by n we will denote the number of network nodes and by Ω the set of all possible $\binom{n}{2}$ edges (i.e. pairwise interactions) between these nodes.

- Model $\mathcal{G}_{n,m}$: select the m edges of G by selecting them uniformly at random, independently of one another from Ω.

- Model $\mathcal{G}_{n,p}$: include each edge of Ω in G independently of the others and with probability p.

- Model $\mathcal{G}_{n,R_0,d}$: generate n points in some d-dimensional metric space uniformly at random and draw an edge between two points only if their distance is at most R_0. This is the *fixed radius model*.

- Model $G_{n,m,p}$: each node i of the n available creates a set S_i by selecting uniformly at random each of the available m objects with probability p. Then an edge is formed between two nodes i, j only if $S_i \cap S_j \neq \emptyset$. This is the *random intersection graph model*.

- Model $\mathcal{G}_{n,m,g}$: According to this model, each of n available network nodes selects uniformly and independently of the others to have degree s, where s is either 0 (i.e. the node chooses to be disconnected from the rest of the network) or it ranges from $1 \leq m \leq n-1$, which is the minimum degree other than 0 that is allowed in the network, up to $n-1$. The probabilities with which these choices are made are defined through the following probability density function, where $2 < g < 3$:

$$p(k) = \{ \begin{array}{ll} (g-1)m^{(g-1)}k^{(-g)}, & m \leq k \leq n \\ 0, 1 \leq k < m & \\ \left(\frac{m}{n}\right)^{g-1}, & k = 0. \end{array}$$

We set $P(k) = \mathbf{Pr}[\text{degree} = k] = \int_{t=k}^{k+1} p(t)dt$, with $0 \leq k \leq n-1$. After the degree choices are made, the graph to which the degree sequence corresponds (up to isomorphism) is formed to be the random graph based on the chosen degrees of the nodes.

Theorem. 1 *(Arratia & Liggett, 2005, case(d)) Let* D *be a positive integer valued random variable and let* $D_1, D_2, \ldots, D_n$ *be an i.i.d. (independent, identically distributed) sequence of random variables with the distribution of* D*. Let* Λ *be the set of limit points of the sequence of probabilities* $\mathbf{Pr}[(D_1, D_2, \ldots, D_n)$ *is graphical*$]$*. Suppose that* $0 < \mathbf{Pr}[D$ *is even*$] < 1$*. Then the following statement holds (only case (d) of the theorem is stated below):*

If for the expectation of the random variable D, $\mathbf{E}[D]$*, the inequality* $\mathbf{E}[D] < \infty$ *is true or if* $\sup_n n \log n \mathbf{Pr}[D \geq n] < \infty$ *then* $\Lambda = \{\frac{1}{2}\}$*.*

Lemma. *The random variable k with distribution as defined in the model $\mathcal{G}_{n,m,g}$ satisfies the first condition of case (d) of the Theorem of Arratia and Liggett, i.e.* $\mathbf{E}[k] < \infty$.

Proof. We compute $\mathbf{E}[k]$ as follows:

$$
\begin{aligned}
\mathbf{E}[k] &= \int_{t=0}^{n} tp(t)dt \\
&= \int_{t=0}^{1} \left(\frac{m}{n}\right)^{g-1)} dt + \int_{t=m}^{n} t(g-1)m^{g-1}t^{-g}dt \\
&= \left(\frac{m}{n}\right)^{g-1} + \frac{m(g-1)+(g-1)(m^{g-1}n^{2-g})}{g-2}, \qquad (1)
\end{aligned}
$$

which is finite (since $g > 2$) if m does not tend to infinity along with n. $\square$

The alphabet of the first order language of graphs consists of the following:

- Infinite variable symbols, e.g. $z, w, y \ldots$ representing graph vertices.

- The binary relations "$==$" (equality between graph vertices) and "$\sim$" (adjacency of graph vertices) which can relate only variable symbols, e.g. "$x \sim y$" means that the graph vertices represented by the variable symbols x, y are adjacent.

- Universal, $\exists$, and existential, $\forall$, quantifiers (applied only to *singletons* of variable symbols).

- The Boolean connectives used in propositional logic, i.e. $\vee, \wedge, \neg, \Longrightarrow$.

Definition. (Extension statement $A_{r,s}$) *The extension statement $A_{r,s}$, for given values of r, s, states that for all distinct $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$ there exists distinct z adjacent to all y_is but no x_js.*

The importance of the extension statement $A_{r,s}$ lies in the following. When applied to the first order language of graphs, if $A_{r,s}$ (for *all* r, s) holds for a random graph G (in some random graph model) with probability tending to 1 asymptotically with the number of vertices of the graph, then for every statement A written in the first order language of graphs either $\lim_{n\to\infty} \mathbf{Pr}[G(n,p) \text{ has } A] = 0$ or $\lim_{n\to\infty} \mathbf{Pr}[G(n,p) \text{ has } A] = 1$.

Let A_i be the event that for the ith formed set of $s+t$ vertices, with $1 \leq i \leq \binom{s+t}{s}\binom{n}{s+t}$, a vertex z cannot be found that is connected to all s vertices and to none of the t vertices. Then

$$\begin{aligned} \mathbf{Pr}[A_{s,t} \text{ fails in } \mathcal{G}] &= \mathbf{Pr}[A_1 \vee A_2 \ldots A_{\binom{s+t}{s}\binom{n}{s+t}}] \\ &\leq \sum_{i=1}^{\binom{s+t}{s}\binom{n}{s+t}} \mathbf{Pr}[A_i]. \end{aligned} \tag{2}$$

Lemma. *The probability that the extension statement $A_{s,t}$ fails for a random graph of the $\mathcal{G}$ model is bounded from above as follows:*

$$\mathbf{Pr}[A_{s,t} \text{ fails in } \mathcal{G}] \leq \binom{s+t}{s}\binom{n}{s+t}\left[1 - P_e^s(1-P_e)^t\right]^{n-(s+t)}$$

with P_e the probability of an edge $\mathcal{G}$, assuming that the edges appear in the graph independently of each other.

Theorem. 2 *For the random model* $\mathcal{G}_{n,m,g}$ *with* $m = \Omega\left(\frac{n}{\log n}\right)$ *the extension statements hold with probability tending to 1.*

In applying the Lemma to the $\mathcal{G}_{n,m,g}$ random graph model, the difficulty lies in the fact that the probabilities that an edge exists between a vertex and a number of other vertices are not independent because the degrees of vertices are fixed once they are chosen. To see this, consider three vertices z, x, y with degrees d_z, d_x and d_y respectively. Then we consider the probability that an edge exists between vertices z, x (denoted by $z-x$) given that an edge exists between z, y (denoted as $z - y$):

$$
\begin{aligned}
\mathbf{Pr}[z-x|z-y] &= \sum_{d_y=m}^{n-1}\sum_{d_z=m}^{n-1} \mathbf{Pr}[z-x|z-y,d_y,d_z]P(d_y)P(d_z) \\
&= \sum_{d_y=m}^{n-1}\sum_{d_z=m}^{n-1} \mathbf{Pr}[z-x|z-y,d_y,d_z]P(d_y)P(d_z) \\
&= \sum_{d_y=m}^{n-1}\sum_{d_z=m}^{n-1} \left(\frac{d_z-1+d_x}{n-1} - \frac{(d_z-1)d_x}{(n-1)^2}\right) P(d_y)P(d_z) \\
&< \sum_{d_y=m}^{n-1}\sum_{d_z=m}^{n-1} \left(\frac{d_z+d_x}{n-1} - \frac{d_z d_x}{(n-1)^2}\right) P(d_y)P(d_z) = \mathbf{Pr}[z-x].
\end{aligned}
$$

Thus, the events that edges exist between a vertex z and a number of other vertices are *negatively* correlated. In view of this fact, we will proceed as follows:

$$\mathbf{Pr}[A_{s,t} \text{ fails in } \mathcal{G}_{n,m,g}] \leq \binom{s+t}{s}\binom{n}{s+t}$$
$$\exp[-\mathbf{Pr}[z - y_1, z - y_2, \cdots, z - y_s, z \not- x_1, z \not- x_2, \ldots, z \not- x_t](n-(s+t))].$$

The notation "$-$" means "existence of an edge" while the notation "$\not-$" means "absence of an edge".

We set $A = z - y_1, z - y_2, \cdots, z - y_s$ and $B = z \not- x_1, z \not- x_2, \ldots, z \not- x_t$. Then it is easy to see that A, B are positively correlated events, since knowing that z is not connected to any of the vertices $x_1, x_2, \ldots, x_t$ results in an increase of the probability that it is connected to all of the y_is. Thus, $\mathbf{Pr}[AB] > \mathbf{Pr}[A]\mathbf{Pr}[B]$. And since the function e^{-x} is monotone decreasing with x we can write

$$\mathbf{Pr}[A_{s,t} \text{ fails in } \mathcal{G}_{n,m,g}] \leq \binom{s+t}{s}\binom{n}{s+t}$$
$$\exp[-\mathbf{Pr}[z - y_1, z - y_2, \cdots, z - y_s]\mathbf{Pr}[z \not- x_1, z \not- x_2, \ldots, z \not- x_t](n-(s+t))].$$

We will now consider the event A alone. The events whose conjunction A is, are negatively correlated and, thus, we cannot replace $\mathbf{Pr}[z{-}y_1, z{-}y_2, \cdots, z{-}y_s]$ with the product $\mathbf{Pr}[z{-}y_1]\mathbf{Pr}[z{-}y_2]\cdots\mathbf{Pr}[z{-}y_s]$ and, still, have a valid inequality. We will, instead, compute exactly the probability $\mathbf{Pr}[z{-}y_i|z{-}S]$ with $z{-}S$ denoting the event that z is connected to all vertices in $S \subseteq \{y_1, \ldots, y_s\}$ subject to the constraint $y_s \notin S$.

The same remark about negative dependence holds with the events in $\mathbf{Pr}[z \not\vdash x_1, z \not\vdash x_2, \ldots, z \not\vdash x_t]$

This probability is

$$\mathbf{Pr}[z - y_i | z - S] = \sum_{d_y=m}^{n-1} \sum_{d_z=m}^{n-1} \left(\frac{d_z - |S| + d_y}{n-1} - \frac{(d_z - |S|)d_y}{(n-1)^2} \right) P(d_y)P(d_z) \quad (3)$$

because the conditional $z - S$ simply excludes $|S|$ possible vertices that could connect z to y_i. Since $|S| \leq s$, a constant, we can see that the presence of $|S|$ does not significantly affect the result of the computation as n tends to infinity. Thus, we can approximate $\mathbf{Pr}[z - y_1, z - y_2, \cdots, z - y_s]$ with the product $\mathbf{Pr}[z - y_1]\mathbf{Pr}[z - y_2] \cdots \mathbf{Pr}[z - y_s]$, as n tends to infinity.

In the same way, we can replace $\mathbf{Pr}[z \not- x_1, z \not- x_2, \ldots, z \not- x_t]$ with the product of the individual probabilities.

After doing this, we can see that setting m to be $\Omega\left(\frac{n}{\log n}\right)$ gives the desired convergence to 1 of the probability that the extension statements hold

The "catch": the random graph formed with $m = \Omega\left(\frac{n}{\log n}\right)$ may not be realizable since now the degree expectation tends to infinity and, thus, the theorem of Arratia & Liggett cannot prove realizability

We conjecture that the graph is not realizable when $m = \Omega\left(\frac{n}{\log n}\right)$ and, thus, (reversing the argument), realizable $\mathcal{G}_{n,m,g}$ graphs do not exhibit 0/1 behaviour for properties written in the first order language of graphs

Interesting questions

- Define random graph models suitable for modeling DT contexts that have/do not to have 0/1 probability laws for properties defined using the first order language of graphs

- Find conditions for having

$$\mathbf{Pr}[A_{s,t} \text{ fails in } \mathcal{G}] \;=\; \mathbf{Pr}[A_1 \vee A_2 \ldots A_{\binom{s+t}{s}\binom{n}{s+t}}]$$

exhibit, itself, a threshold behaviour. This requires the use of advanced methods for approximating conjunctions/disjunctions of dependent probabilities (e.g. McDiarmid's, Suen's and Janson's inequalities) – we have a "two-level" dependency: in computing $A_i = \mathbf{Pr}[z - y_1, z - y_2, \cdots, z - y_s, z \not- x_1, z \not- x_2, \ldots, z \not- x_t]$ and, then, $\mathbf{Pr}[A_1 \vee A_2 \ldots A_{\binom{s+t}{s}\binom{n}{s+t}}]$

- Define random graph models that allow “non-homogeneity” in the produced graph

- What happens when properties are considered that cannot be written in the first order language of graphs? It can be proved that there exist such properties that do not have a 0/1 behaviour

THANK YOU!