Advances in Networks
2015; 3(2): 8-21
Published online August 3, 2015 (http://www.sciencepublishinggroup.com/j/net)
doi: 10.11648/j.net.20150302.11
ISSN: 2326-9766 (Print); ISSN: 2326-9782 (Online)

Predicting Behavioural Evolution on a Graph-Based Model

Arnold Adimabua Ojugo[1], Rume Elizabeth Yoro[2], Andrew Okonji Eboka[3], Mary Oluwatoyin Yerokun[3], Christiana Nneamaka Anujeonye[3], Fidelia Ngozi Efozia[4]

[1]Dept. of Math/Computer, Federal University of Petroleum Resources Effurun, Delta State, Nigeria
[2]Dept. of Computer Science, Delta State Polytechnic, Ogwashi-Uku, Delta State, Nigeria
[3]Dept. of Computer Sci. Education, Federal College of Education (Technical), Asaba, Delta State, Nigeria
[4]Prototype Engineering Development Institute, Fed. Ministry of Science Technology, Osun State, Nigeria

Email address:
arnoldojugo@yahoo.com (A. A. Ojugo), ojugo_arnold@yahoo.com (A. A. Ojugo), ojugo.arnold@fupre.edu.ng (A. A. Ojugo), rumerisky@yahoo.com (R. E. Yoro), agapenexus@hotmail.co.uk (M. O. Yerokun), an_drey2k@yahoo.com (A. O. Eboka), anujeonyechristy@gmail.com (C. N. Anujeonye), fenngo31@yahoo.com (F. N. Efozia)

To cite this article:
Arnold Adimabua Ojugo, Rume Elizabeth Yoro, Andrew Okonji Eboka, Mary Oluwatoyin Yerokun, Christiana Nneamaka Anujeonye, Fidelia Ngozi Efozia. Predicting Behavioural Evolution on a Graph-Based Model. *Advances in Networks*. Vol. 3, No. 2, 2015, pp. 8-21.
doi: 10.11648/j.net.20150302.11

Abstract: Corruption is the bane of any economy. Its malady cuts across religious, socio-economic and political system of Nigeria. With a fast and contagious spread through the nation's socio-economic and political strata, its adverse malignant effect is today, difficult to treat. This study models its contagion via an agent-based graph-diffusion model. Graphs are now quickly becoming the dominant life-form of most activities in a society, with human actors as nodes. Actors have ties that bind them to others via interaction as they form a social graph that analyzes the agent's local feats via interaction to impact on the society as a global structure. Study explores the graph's rich connective patterns and personal-networks as actors influence each other, so that graph's behaviour evolves to orchestrate a relationship in probabilities of observed data and recognize patterns that aid decision making via its convergence to predict the expected number of final adopters as its optimal solution in a multi-peak function.

Keywords: Stochastic, Immunize, Network, Vertices, SIS, SIR, Function, Search Space, Solution, Models

1. Introduction

Many studies have defined corruption from its various frontiers with a degree of disposition that seems quite interwoven. In its simplest form, Corruption is abuse of public office for private gain. Akindele (1995) note it as any form of reciprocal behavior or transaction where a power/office holder can initiate inducement of each other by some rewards to grant (illegal) favour against the ethics and interest of a specific society. Corruption covers range of act: (a) use one's office for pecuniary benefits, (b) gratification, (c) influence peddle, insincerity and illicit gain via an office, (d) full-day pay for less work, (e) slovenliness and tardiness. Osoba (1998) note corruption is an anti-social behaviour that confers improper benefits contrary to legal norms, and undermines the authorities that tends to improve the living conditions of the people.

It occurs in many forms and continues to contribute much immensely to the poverty and misery of a large segment of many nations' populace. Corruption's many adverse effects have led to the craze/crave for wealth accumulation and/or acquisition in recent times (and in Nigeria) where many of such perpetrators when caught, are often celebrated. Aluko (2002) notes that corruption is a social malaise that appears to have become a norm and institutionalized in the Nigerian polity – as it has permeated into every realm of culture and her value-system, as today's young Nigerians are born into it, grow up, live with it and possibly die in it. The aged are not left out as they are re-socialized and begin to conform to its many norms. He notes its contagious nature with a malignant effect to physiology of her political system – such that once set into a facet of the society, it automatically contaminates all strata of system's socio-political structures in ways symmetrical to the spread of virus.

1.1. Nature and Characteristics of Corruption

The upsurge of corruption in Nigeria seems has defied all treatment. Its damage as quantified to the nation's life, its fabric of existence and the culture on which Nigeria is built, has become quite alarming. Corruption, though a global event, the menace has led to the slow progression of daily activities that allows for growth in any socio-economic and political society (Dike, 2011). Corruption continues to enrich those in power; and, strangles the poor and suffering masses – relegating them into abject poverty and hopelessness. It is common belief that corruption in Nigeria is endemic. It affects both government and non-government settings with little to be done about it. Thus, Taylor (2010) amongst other researchers have taken a holistic (broader) view to define the nature of corruption under these types:

a. Political – if politicians/decision-makers who formulate, establish and implement laws in a state, are themselves corrupt. Also, when policy formulation and legislation are tailored to benefit politicians and legislators.
b. Bureaucratic – occurs in the public administration or the implementation end of politics. It is encountered by the citizens daily at places like hospitals, schools, police station, church etc. – and thus, occurs when one obtains business from public sector via inappropriate procedure.
c. Electoral – occurs when votes are bought with money, promises of office special grants/favours, intimidation, coercion and interference with freedom of election.

Other forms of corruption include:

d. Bribery – payment (in money or kind) taken or given in a corrupt relationship such as kickbacks, gratuities, pay-offs, sweeteners, greasing palms scratching back etc.
e. Fraud – involves tricks, deceit counterfeiting, racketing, smuggling, swindling and forgery etc.
f. Embezzlement – is theft of public resources by public officials. It is when an official of the state steals from the public institution in which he/she is employed. It is the most common means of wealth accumulation in Nigeria due to lack of strict regulatory systems.
g. Extortion – is money and other resources extracted via coercion, violence or threats to use force. It is often seen as extraction from below.
h. Favoritism – is the mechanism of power abuse implying a highly biased distribution of state resources. However, many see this as a natural human proclivity to favour friends, family, and anybody close and trusted.
i. Nepotism – is a special favoritism in which an office holder prefers family members as well as exempts them from the application of certain laws or give them undue preference in allocation of scarce resources.

1.2. Causes of Corruption

Causes of corruption are myriad – with both political and cultural variables. Evidence links corruption to social divide, ethno-linguistic divide and religious diversity. Others have attributed its widespread to non-democratic societies as well as societies branded as neo-patrimonial, kieptocratic and prebendal. It implies that the political system and the culture of a society can make its citizens more prone to corrupt activities. Some other fundamental factors that engender corruption in nations like Nigeria are: (a) great inequality in distribution of her wealth, (b) political office as the primary means of gaining access to wealth, (c) conflict between changing moral codes; (d) non-robust enforcement of laws by societal mechanism, (e) absence of a strong sense of national community, (f) great obsession with materialism, and (g) compulsion for a shunting to affluence, glorification, and approbation (of ill-gotten wealth) by the general public office holders.

1.3. Corruption as a Function of Behavioral Evolution

Corruption has become a way of life in almost every facet of the Nigerian nation/society – where office holders abuse the powers of such office for their own personal gains. Their subordination (not left out in the whole act) act as a supply chain cum network to partake of the crumbs (gains) as the act is perpetrated, and are thus – viewed as partners. Several theories serve to explain and understand collaboration in the context of networks, graphs and partnership. It emanates from theoretical framework that consists of alliances formed via ties, transaction costs as a function of network structure and social exchange that determines threshold value of each actor within the graph (social system/society). The key feat is the exchange of relationships (dyads/ties) between the office holders and those seeking unmerited favours in such exchange by drawing largely on the cross-sectional data as well as tracks how broader patterns of collaboration can be formed over time (Akkermans and Romme, 2008).

The nature of the graph network will draw on notions of both parties disposition. Disposition draws from concepts such as opportunism, trust, flexibility and learning as the key drivers of tie-strength and actual partnership behaviour as embedded within the parties and society as transactions occur. Thus, partnership in itself is quite ambiguous due to tension between trust and power as the parties' relationship ensues (Ojugo et al, 2014). Such a complex graph, understood in terms in dyadic game involving actors A and B, in which an actor A reacts to the actor B's action in a certain way. Even in games involving several actors, it may be impossible to derive the attitude of actors from behaviour because the interactions are quite too complex (Elias, 1998). He further notes inner motivations and dispositions (attitude) cannot be derived from the outward behaviour of actors – be they individuals, groups and societies due to social figuration, which help to reflect and determine such actor's behaviour. He describes this further as a game in which the move made by an actor determines the character of the actor who made such move – as any of the explanations may be justifiable but never sufficient. A move can no longer be explained in as short, unilinear casual sequences – nor is it based on just an actor's character. Rather, each move in the game as played can only be interpreted in the light of how previous and preceding moves, made by both players are intertwined, and of the specific figuration and stratifications, which has resulted from this intertwining (Elias, 1998).

The issue is reinforced as the process consists of a large number of actors – as long-term partnership behaviour has little of an actor's actual disposition at any given time. Since such collaborative process involves many actors, who are characterized by ambiguity, enactment and retrospective decision making – study is more concerned with emergent processes and feats, attuned to sequences that unfold in such generative settings as amplified via small events with large consequences. Such small beginnings generate unanticipated consequences via complexity theories and do not remain small in time (as diffusion model). They change in size – to constrain other events and spread through what others reify into groups, institutions and societies (Weick, 2004).

2. The Actor-Based Theoretical Framework

Social graph consists of ties that bind actors together via a model to help analyze social entities along various theories to explain the observed patterns within (Wasserman and Faust, 1994). It propagates local feats of actors that emerge as global patterns, examine various dynamics in relationship between actors and locates influential entities within – to result in social connection convergence between relational contacts theoretically (Valente, 1996).

Social graphs use two concepts: (a) *realist,* where social actors (family/colleague) with behavioural, psychic, verbal and emotional ties in a multi-dimension contact system have such awareness shared by members. The group generates its boundaries, norms and rules that must guide each member as they perform functional role towards each other in the achievement of a common goals, and (b) *nominalist* – an actor imposes a defined boundary on its members to identify group (Marsden and Campbell, 1990; Granovetter, 1985). A social graph/network is endowed with the potential of being transformed into a completely realist group so long there is enough interaction; Also, can be both nominalist and realist (Wasserman and Faust, 1994).

Visual representation helps convey results of explored nodal-ties and complex feats captured via signed structural quantities that represent node relationships. Positive edges for positive relations (marriage, friendship, colleagues etc); while negative edges are for negative relation (hatred, anger etc). If A→B and B→C are positive; C→A is a negative – it is an *unbalanced* cycle. Also, if B→A is positive, and both A→C and B→C are negative – these groups will be likely morphed into *balanced* one. Such concept is used to predict tie-behavioral evolution for both *balanced* (where actors will unlikely change their opinions about other actors in same group) and an *unbalanced* (actors will likely change opinions) cycles (Granovetter, 1985).

A powerful role of networks is to bridge local feats as it blossoms into global patterns – explaining how simple node and ties process impacts a complex effect that ripple through a population system. Each node shapes the graph's evolution in time, adapting themselves to varying forms as need arises, which defines tie-strengths. Graph is denoted as G = (V,E). Each node $i \in$ V has set of ties $m \in$ E that is either self-linked (loop), single-linked or multi-link (with undirected or directed). Each node has a corresponding set of actors to which they are either linked or isolated from. The links are either weak or strong to note the social relationship between such nodes, which are measured via dyads *D*.

Social graphs aim at two purposes namely: (a) to better understand how networks evolve, and (b) study dependent social processes like innovation diffusion and data retrieval via models to specify how local interaction of agents' feats are explored to a global pattern.

2.1. Agent Based Modeling: An Overview

The future continually bewilder and leave us curious as we seek to control on all and every frontier, most of our daily activities and dealings in our society. Such knowledge and insight is achieved to help us plan ahead future conditions of our daily activities and to know-in-time, the scale of other dependent events as a lot of what we do relies on it. We may inquire to know how office power and privileges can help us amass wealth for our nuclear and extended family – due to cravings to belong to the affluent in the society; whereas our immediate society is more concerned about impact of such wealth and how it was amassed, to the nation. Such events predicted, help experts make informed decisions about the future from observed values, made manifest via models. Models thus, are tools that yield knowledge statement of the future we sought, to provide real-time prediction that helps us plan in time magnitude of a probable event as it reaches maximum (Ojugo et al, 2012). Its reliability can be questioned as its results are seen as prophesies rather than prediction (Macy and Willer, 2002).

A flock of bird flying in tight formation will collectively form an image with a goal movement as a single organism. Yet the flock choreographing in such grace has no group-leader bird. Instead, each bird reacts to the movement of its immediate neighbors, to result in hypnotic patterned rhythm and highly-nonlinear. Modeling such elegance not governed by any system – is tedious and difficult due to its dynamism, complexity and nonlinear nature. But, it can be modeled as an aggregation of local feats interactions via 3-simple rules: (a) separation – each bird does not get too close to another, (b) alignment – each bird matches its direction and speed to nearest bird, and (c) cohesion – each bird stores in memory the perceived center of the flock and its immediate neighbor (Reynolds, 1987). It models each bird as an agent with local feats interaction to yield a highly realistic flight formation via simple rules – to result in the theory of Agent Based Modeling – ABM (Ojugo et al, 2014).

ABM, applied in graph-design with actors of highly self-organized, nonlinear, path-dependent, dynamic and complex graph is best understood as emergent local feats interacting between adaptive actors in response to varying external and internal influences received – and results in a global pattern (Epstein and Axtel, 1996). Studies have harnessed the many potential of ABM as tools in relational method modeling with its fundamental focus on emergent social structure and social order via local interaction. It posits: (a) a theoretical frame of dynamic social graphs shaped via interactions of actor, and (b) tests social-learning theories that manipulate feats of the structure such as its topology, stratification etc (Simon, 1998; Kaufman, 1996; Macy and Willer, 2002).

2.2. Diffusion Immunization Model Framework

There are two major susceptible-infect (SI) models namely: Susceptible-Infect-Remove (SIR) and Susceptible-Infect-Susceptible (SIS). In SIR, a node may be in any of these states: (a) susceptible: if the node is not exposed but will soon be, (b) active: if the node has been exposed and has adopted the innovation as well as influence others to do same, and (c) removed: if the node had been exposed to corruption and has been recovered. The node is permanently immunized and can no longer participate in propagation – so that such a node cannot be exposed twice to corruption. Conversely in SIS, a node can be cured but not immunized. Thus, it can be exposed again. Such node switches between susceptible and immunized.

Giakkoupis et al (2010) and Lahiri and Cebrian (2010) A graph holds these definitions as true:

a. A graph is either directed or undirected, with nodes as actors as v ∈ V; and ties (u,v) ∈ E that allows for interactions between two/more actors in the system. It also assume that the graph is drawn from a specific family (algorithm consider all possible graphs). For G = (V,E) as a dynamic network, E is set of edges that are time-stamped, $(u,v)_t \in E$ are interactions at $t \in Z^+$. In a typical SI setting, nodes are initialized as *active* – so that in discrete time-steps and at each time evolution, an active actor/node is exposed to the innovation for adoption. This continues till a stop criterion is satisfied or there are no more inactive vertices.
b. Immunization model aims to maximize spread of anti-corruption practices and stay immunized so it cannot engage in corruption no more; And, is conceptually removed from G. Cost of immunization model are immunized nodes.
c. Knowledge of the propagation model will help us plant *d* copies of seed-set nodes that have adopted the anti-corruption practices in the network so as to maximize speed of its spread given by F_d. An adaptive seed-set has knowledge of choices made by the immunization algorithm; while a randomized seed-set simply places uniform cum random copies of seed-set on the network.

The susceptible-infect model has two basic types namely:

a. Susceptible Infect Remove Model – At discrete-time at t = 0, the graph is inserted with actors exposed to the innovation – so that if an actor x is exposed and has adopted the innovation at time t, it has single chance to convince its immediate neighbour y that is currently active but yet-to-be-exposed. The probability that x succeeds with y is P_{xy}. If x succeeds, y is exposed and has adopted the innovation at time at t+1; Else, x tries again in the future (even if y adopts the innovation via another actor's personal network or neighbour). This process continues and stops after *n*-steps (at time t – 1) – at which point there are no more active actors to be exposed to the innovation. It requires an actor to be exposed exactly once (Kempe et al, 2003). The graph can be of size M, with M_d subset of nodes and d copies of seed-set actor nodes placed on the network. With propagation complete, $S(M_d,G)$ is the expected number of final adopters. Expectation can exceeds all random choices depending on the diffusion model and nature of graph structure in use. Eq. 1 describes the maximum expected number of final adopters, which exceeds all possible initial seed-set choice adopter placements.

$$S_d(G) = \max_{M_d} S(M_j, G) \tag{1}$$

The subset $A_d = arg\max_{M_d} S(M_d, G)$ corresponds to choices made by an adaptive adversary. $S_d(G)$ is epidemic spread in G and a similar definition of epidemic spread of randomize adversary as in Eq. 2 in which case, define it to be the expected epidemic spread where the expectation takes over all possible positions of the d viruses placed on the network and given by:

$$S'_d(G) = \mathrm{E}_{\mathrm{M_d}}[S(M_d, G)] \tag{2}$$

b. Susceptible Infect Susceptible – At discrete time-step at t = 0, the graph is inserted with actors exposed to the innovation – so that if an actor x is exposed and has adopted the innovation at time t, it has single chance to convince its immediate neighbour y that is currently active but yet-to-be-exposed. Probability x succeeds with y is P_{xy}. If x succeeds, y is exposed and has adopted the innovation at time at t+1; Else, x tries again in the future (even if y adopts via another actor's neighbour). Also, actor y can decide to reject the innovation at another time-step (due to path dependence, morphing of ties, external influences and other feats). This process continues and stops after *n*-steps (t – 1), at which point there are no more actors to be exposed to innovation. It requires an actor to be exposed more than once (Acemoglu et al, 2012). The graph can be of size M, with M_d subset of nodes and d copies of seed-set actor nodes placed on the network. With propagation complete, $S(M_d,G)$ is the expected number of final adopters. Thus, the process continues to evolve in time to either propagate or eventually die. An actor x adopts based on the rate of $\frac{\beta}{\delta}$ and probability β. At same time, an actor may reject the adoption after having previously adopted probability δ. With an adjacency matrix T, $\lambda_1(T)$ is largest eigen-value of T. Thus, $\frac{\beta}{\delta} < \frac{1}{\lambda_1(T)}$ is true as threshold and is sufficient for quick recovery as easily proven (Ganesh et al, 2005; Wang et al, 2003).

Anti-corruption practices are best seen as an immunization problem, no matter the diffusion model used – and for this study, we adopt SIR immunization model. While estimating the parameters of a social graph in diffusion, 3-basic criteria must be addressed as model converges namely: (i) *Targeting* (which actors are targeted as *initiators* to result in maximum adoption of innovation, irrespective of the diffusion model used with existence of *seed* set, how does the innovation penetrate highly clustered and cohesive set with no seed set?), (ii) *Extent* (if *seed-set* of initially exposed actors exist, what is expected number of final adopters given a certain time?) and (iii) *Blocking* (which actors must be targeted to maximize adoption of innovation by the expected number of adopters within a certain time function so that they remain and stay immunized?).

The expected number of final adopters of the innovation – is a function of diffusion model's convergence in time with respect to tie strengths, network structure, cohesion and the seed set with a member in agent's personal network for the diffusion. Gilbert and Kalahalois (2009) Social graphs can have a defined set of parameters that helps it to engage in estimation and prediction as the graph aims at convergence of the final number of adopters on a time varying network.

Graphs have two parameters groups:

a. *Dependent, Local-Emergent* parameters are those feats or properties whose values are tied to the nature of task. They consist: (i) *n* number of agents (nodes), (ii) *m* number of ties (edges) (iii) *tie-strength* (encoded as set of predictive variables, error term, pairwise interactions between predictive variables etc), (iv) *agent's threshold value*, (v) *path-dependence* (system internal influences), (vi) *agent's personal network* (P_iG) and (vii) *seed-set* choice and/or selection.
b. Conversely, *Independent* parameters have their values tied as both external influences and resultant expected number of final adopters to yield a global pattern for the agents (both as personally and as clusters/community) as they consider adoption or rejection of the innovation. These, we can refer to as: (i) *external influences* that are captured within tie-strength and (ii) *network structure* (encoded within feats such as probability distribution of graph, seed-set, reciprocity, graph radius and diameter, clustering coefficient amongst others).

The orchestrated relationship found as a measure of their correlation from (regressed values as the model goes back in time from the observed dataset, checking the worst or more primitive condition of such observed data-states as well as aim to progress in evolution so that the model will yield future or predicted states in time forward, from the observed data, whose error relationships are found via mean relative error (MRE), mean absolute error (MAE), mean square error (MSE), coefficient of efficiency (COE) and of determination R2 respectively.

2.3. Tie-Strengths

Events modeled as social graphs will never treat all actors as equal. Thus, it distinguishes between trusted friends from total strangers (in that continuum) via relationship spectrum as provided by *tie strengths*. Loose acquaintances that an actor interacts with on demand are said to have *weak ties* with between such actors. Though weak, they can help such actors generate ideas and expedite knowledge transfer across groups. Conversely, interaction between trusted friends and family are said to be *stronger-ties*, which affects the actor's emotional health and can lead a society to safe heavens in time of crisis. Whether a link between actors exists or not, relationships have their own few properties that helps define tie-strength that exists between agents as they form clusters, cliques and communities (Newman, 2003b).

Dyads are all pair of interactions to measure relationships of *n* actors with *m* ties – resulting in an m×n binary matrix of elements in G. The ordered pair (i, j) is 1 if it is sampled and 0 if not. Both directed/undirected graphs have a set of observed dyads incident in at least one sampled node.

Tie-strengths have 7-dimensions that manifest in various forms to include: (a) time structure, (b) emotional intensity, (c) mutual intimacy and (d) reciprocal services (Granovetter, 1974). Burt (2004) note structural variables and factors like (e) topology and (f) emotional support from informal circles indicate a stronger tie. Lin et al (1981) note that: (g) social distance as embodied in socio-economic status, education, race, political affiliation and gender of an actor – all which influences tie-strength. In practice, the structural variables are substituted with simple proxies such as communication recency (Lin et al, 1978), interaction reciprocity (Friedkin, 1980), mutual friendship (Shi et al, 2007) and frequency of the interaction (Gilbert et al, 2008).

Gilbert and Karahalois (2009) aimed to precisely unpack tie-strength predictors in which their participants overcame issue of retrospective informant accuracy as they harnessed the benefits of a long friend's lists with the rich interaction histories of social media. Ojugo et al (2014) extended the work of Gilbert and Karahalois 2009 with the potential of a feedback into the society in ways that benefit users as in Eq. 3. Thus, tie-strength is modeled as a linear combiner, where R_i represents number of predictive variables used in task, e_i is error term, D_i is dyads pairs, N_i is network structure and EI_i are external influences.

$$S_i = \propto\ + \beta R_i + \gamma D_i + N_i + EI_i + \ e_i \qquad (3)$$

2.4. Network Structure N_i

Acemolgu et al (2012) A cluster is a strong candidate of a cohesive set or clique by itself due to large number of ties amongst members. As clusters increase, actors form more cliques with others in close proximity, which also decreases the set cardinality and increase the number of sets in graph. High-clustered graphs have short-path length, large number of cliques with small cardinality. It results in large expected number of final adopters, introduces a close knit relationship between clustering coefficient and bound for number of final adopters. Such graphs are quite beneficial for complex diffusion process (contagion) to reinforce adoption of the innovation as they are more likely to be exposed to multiple adopters and overlapping influences via such short path-lengths during diffusion. Its merit is, in the existence of a seed set adopter inside the cluster. In contrast, with no *seed-set* adopter in a locally dense cluster, it is highly stable and may resist adoption of innovation. Thus, while clustering reinforces adoption if the innovation penetrates, it also weakens adoption by making penetration more difficult for small cluster coefficient *k*. Graphs with smaller clusters have long path-lengths to diffuse innovation more and further (Gilbert and Karahalois, 2009; Acemoglu et al, 2012).

N_i encodes an idea that its structure is dependent on the structural predictive variable in tie-strengths with respect to: (a) tie history, (b) reciprocity, and (c) cohesive clusters and its coefficients. A major reason in using clustering includes: (i) there is a direct relationship between cohesiveness and clustering, and (ii) clustering has been used extensively in many studies to capture graph network structure as it also allows for easy comparison of result (Acemoglu et al, 2012), which is given as thus:

$$N_i(G) = P_{\theta,G} + \lambda_0 \mu_L + \lambda_1 Med_L + \sum_{t=0}^{t-1} \sum_{i \in L} \lambda_t (s - \mu_L)^t + \lambda_5 Min_L + \lambda_6 Max_L \quad (4)$$

where L = S_j and i,j are mutual friends. N_i uses the following parameters: (a) $P_{\theta,G}$ to encode the system/graph's probability distribution of actors via defined upper and lower bound of agent dispositions in the system, (b) μ and Med are actor's cum system threshold respectively of graph, (c) expression $\sum_{t=0}^{t-1} \sum_{i \in L} \lambda_t (s - \mu_L)^t$ encodes learning with momentum as a function of its convergence in time/outcome of the diffusion process, and (d) Max/Min values are upper/lower bounds of final adopters.

Since each node *i's* neighbour has potentially unique set of mutual friends, the model uses 5-descriptors of tie-strength distribution to describe the graph's structure thus: (a) mean number of ties to cater for cohesion among the nodes of the graph, (b) variance in the number of ties to cater for nodal and joint degree of separation in the graph, (c) kurtosis of the ties to cater for clusters and its coefficient, (d) minimum number of ties in the graph, which caters for local density and degree of distribution, and (e) maximum number of ties in graph, which caters for reciprocity – all belonging to the structural dimension of ties as associated to clustering and cohesiveness of the graph system; while it also introduces a dependency that tie-strength also depends on other tie-strengths (Gilbert and Karahalois, 2009).

2.5. Threshold Models and Path Dependence

A social graph is modeled as tie-patterns that bind loosely connected actors, as required to impact a proposed learning outcome via diffusion model for an innovation in a social system (Granovetter, 1978). It provides a structure in which actors form their opinions of the introduced innovation as well as reflects how actors' interactions and local feats can influence the consideration of adoption or its consequent rejection (Newman, 2003a; Rice, 1994). As ties are formed, actors vary from one another – in their willingness to risk in adoption of a new innovation, product, ideas and behaviour. Thus, most actors are quite reluctant as they will rather wait until others have adopted such innovation first. The measure of this willingness to adopt is termed *threshold value* of an actor for the innovation (Valente, 1996).

Threshold is collective behaviour where an actor considers adoption based on the proportion of mutual friends that have already done so in the social system (Granovetter, 1978). An actor thus, adopts based on the behaviour of other agents in his immediate neighbourhood (personal network of actors that can influence him to adopt). Actor of low-threshold for a specific innovation will engage before many others; those of high-threshold only engage after many have done so; while actors of same threshold adopts at different times depending on the nature of influences received (Valente, 1996; Marsden and Campbell, l990).

Acemoglu et al (2012) Stochastic threshold of agent *i* from a linear threshold model (Kempe et al, 2005) and collective point of diffusion is: If M = {1,2,3,…N} agents of a graph has V-vertices, E-edges and w-weighted probabilities for its nodal feats denoted by G(V,E,w). With no self-loops in graph, (i,j)∈E is directed from i to j and its weight w(i,j) = 0 only if (i,j)∉E. w(i,j) = [0,1] as $\sum_{i \in J} w(i,j) \leq 1$. Thus, an actor's personal network has a neighbour set of the agent i∈M given by Pi(G) = {j | (j,i) ∈ ε} to consist of agents who can potentially influence agent *i* in G.

If at iteration time t = {0,1,2,…,n), a subset of agents Φ(0) ⊆ M is selected as *seed* set, known as group of innovators exposed with threshold randomly uniform between [0.5,1] to reflect our lack of knowledge of the agent's true threshold (see Kempe et al, 2005). If the *seed* set adopts innovation at t = 0, at t > 0 – actor i ∈ M \ Φ(0) will adopt innovation if at least $\phi_i \in \Phi$ [0,1] fraction of the agent's neighbors are in the seed set. and the expected number of final adopters are given by Eq. 5:

$$FA_d = \frac{|\Phi(0) \cap P_i(G)|}{|P_i(G)|} \geq \phi_i \Rightarrow i \in \Phi([0,1]) \quad (5)$$

Set $\Phi([0,1])$ are actors un-exposed; But, who will consider adoption as persuaded by their personal network. At t ≥ 0, we generalized that: A node i∈M \ $\cup_{l=0}^{t-1} \Phi(l)$will adopt at *t* of linear threshold of Eq. 6.

$$FA_d = \frac{|\{\cup_{l=0}^{t-1} \Phi(l)\} \cap P_i(G)|}{|P_i(G)|} \geq \phi_i \Rightarrow i \in \Phi(t) \quad (6)$$

It is quite powerful to capture the role of backstage actors and interpersonal influences in adoption – with insight into the dynamic, complex relationship in network connectivity with respect to its seed set and threshold values. It fails to capture path dependence amongst other feats/parameters.

Acemoglu et al (2012) Path dependence is the idea that a few minor shocks, insignificant events along the way within the network can alter history's course. Diffusion has been argued to be an extremely fragile task with respect to such small shocks, which in turn goes to imply that the diffusion of innovation process is highly, path-dependent. Two ideas of similar feats, functions, seeding strategy and perceived quality may diffuse differently on the same network due to the different realizations of these minor shocks.

To account for tie-strength and external influences as other feats, we extend Acemoglu et al (2012). Ties help include an actor as member of another actor's personal network with a *seed* set – to allow for faster diffusion and yield cohesive clusters. External influences are the factors that influence an actor's opinion to consider innovation that are not from within the social system – to alter course of history. These include cosmopolitan actions, media influence, campaign, parallel innovation and initiatives etc – as reflected in the model (Valente, 1996).

If we denote the state of an agent at iteration t as $x_i(t)$, so an actor i takes one of 3-possible values: {0,1,-1} as not-yet-adopted, adopted and rejected respectively. At t = 0, the seed set of agents $\Phi(0) \subseteq M$ selected will adopts innovation. Thus, at t > 0, agent $i \in M \setminus \{\cup_{l=0}^{t-1} \Phi(l)\}$ will adopt if atleast $\phi_i \in [0,1]$ fraction of the agent in agent's personal 's network are members of *seed* set based on Bernoulli trial probability $p \in [0,1]$. For each $i \in \Phi(1)$, $x_i(1) = 1$ with p and $x_i(1) = -1$ with probability $1 - p$. Feat p (or FA_d) determines likelihood of adoption conditioned upon its consideration and exposure – so that the larger the value of p, more likely an agent will adopt the innovation when such innovation is considered. Thus, set of agents who have adopted at $t = 1$ forms two sets given as: $A(l) = \{i \in V | i \in \Phi(1), x_i(1) = 1\}$, comprise of actors *seed* set and those that have adopted; while, the other set: $R(l) = \{i \in V | i \in \Phi(1), x_i(1) = -1\}$ comprise actors that are considering rejection. We generalize that at t ≥ 0, expected number of final adopters from modified stochastic threshold model that take into account path dependence, tie strength, and external influences is given by:

$$FA_d = \frac{|\{A(l) \cup_{l=0}^{t-1} \cup \Phi(o)\} \cap S_i(P_i(G))|}{|P_i(G)|} \geq \phi_i \Rightarrow i \in \Phi(t) \quad (7)$$

An agent adopts if: (a) a member of his network has been exposed, (b) members of his network forms a cohesive set, (c) where members of his network is not a cohesive set, then such members that has adopted are mavens, and (d) if one initializes an innovation from a set $\Phi(t)^*$ whose complement is a cohesive set, then such innovation will not be adopted.

Main difference between linear and stochastic threshold model is that an agent does not necessarily adopt if fraction of its neighbours that have adopted is above their threshold. At p < 1, agent can reject with a non-zero probability (due to minor shock and external influences); while at p > 1, agent can accept adoption also as case is (Acemoglu et al, 2012).

3. Materials and Methods

3.1. Data Gathering

The diffusion study uses as its innovation – anti-corruption practices. The study initiative will sought a convergence in time of final number of adopters for youth – with these goals: (a) as youths participate in the scheme, they imbibe idea that corruption is an illegality that must be expunged from the society, and (b) that they are the right machinery needed to steer such new vehicle to redirect Nigeria.

The initiative is designed to convene youths; while having within the 4-day function, a two (2) days lecture scheduled to educated the youths as well as allow them to share their opinion and views for a hostile change of ideology via such anti-corrupt practices. The programme was also designed to allow these youths to choose leaders amongst them, provide small scale loans and educate them on the benefits accrued from hardwork. Year in which youths were first introduced to the concept was known as the time-of-adoption; while the year in which the project was to end noted as t = n; and, the year before that is t = t – 1. Network data were collected by asking these youths to accomplish the following tasks:

a. List their ten (10) best friends within the age range of 11-24years for the project. The timeline of 18months is chosen. Project is based on: (i) implant an idea within actors involved in such transactions grows in time and is quite difficult in that they get to put aside selfish and self-centered gains, (ii) seek in time the convergence of a final number of adopters, and (iii) 18months timeline will only give time that allows for sufficient diffusion of the innovation and project.

Such changes in time in an agent shun such practices as the agents also will see themselves as participants and not as tools to be used – as they build their personal networks and clusters, which in turn will be tested over time. The idea as planted also needs time to be nurtured and groomed in the minds of these agents – bearing in mind that that age range is as a result of the fact that youth are the future leaders of this nation – such that as they graduate, they begin to occupy such office, where powers are exerted.

b. Ten (10) most influential youths in the age range in their immediate community and beyond. The beyond part would also help these agents to reach out to youths outside their immediate domain with the same idea.
c. Ten (10) most influential youths for information spread, who are already involved in anti-corruption training and campaigning process.
d. Ten (10) best youths to organize a cooperative project.

Table 1. *Pre-Analysed Dataset on Social Network.*

Network Feats	**Youth Empowerment as the Innovation**
Number of Communities	8
Clusters in each community	30wards (clusters)
Targeted Number in Age range	25000 youth
Seed Sets Phase	5seed-set nodes per Ward
Time of Diffusion / Year	18months
Graph Probability Distribution	$\mu = 0.5346$ and $\delta = 0.34$
Average time of Adoption	6moths
Lowest and Highest Saturation (final number of adopters)	36% and 80% at t > 0 and t = t – 1 respectively

Time-of-adoption data is based on project initiation date, and is not subject to any error (Coughenour. 1965). Some participants stated their adoption time as earlier as 3years prior to the project initiation in Lagos. While, such data is potentially accurate, its recall event may be erroneous. It is hoped that recall errors are normally distributed. Data measured cosmopolitan external influence by the number of visits to the nearest large city within the past month.

The study does not wish to discuss the effect of the network structure on its convergence time nor the effect of clustering on the diffusion model cum process. Many studies note that innovations diffuse faster on highly clustered networks and much slowly on graphs with small degree of clustering. Our experimental model, rather aims at the expected number of final adopters as a function of the convergence task for a specified duration, the effect of tie-strengths prediction and threshold on actors' behavioural evolution and disposition towards the anti-corruption practices as the innovation.

3.2. Rationale for Using ABM on Social Graph

ABM is used to partner behaviour for the reasons (Akkerman and Romme, 2008; Ojugo et al, 2014):

a. The need to collect longitudinal data on entire network over a period of time complicates empirical studies, and particularly for such complex patterns of interactions in these small world-graphs.
b. Empirical investigating preferences, disposition and other feats for corruption and its consequences, are quite tedious and difficult.
c. People may not know why they do or have done things (taken some decision). Society representatives may not know why the society they are part of, behave in the particular way, and they too may also be reluctant to reveal their true motivations and intentions.
d. Previous studies suggest that emergent feats from local interactions can or may be quite biased. ABM potentials fully harnessed, can model the complex properties of such social systems via the analysis of data generated.
e. ABM modeled on graphs with diffusion of innovation assumes certain dispositions for a number of actors and then observes patterns as they emerge from interactions between such actors.
f. ABM involves elaborate thought experiment to learn about complex adaptive systems rather than build valid representation of the real-world system (Axelrod, 1997; Holland, 1995). Thus, the modeling approach serves to discover unexpected consequences of the interactions in themselves of such simple processes.

The experiment seeks to explore effects of anti-corruption practices in a number of actors as well as find the expected number of final adopters via network parameters with the actors' internal decision rules and position in the graph as they locally interact over time as exposed in the graph-based diffusion model. As such, the study is more interested in the local emergent feats emanating from large-scale effects of such interacting actors of the entire supply network.

3.3. Threshold Adopter Categories

Valente (1996) Major categorization of adopters based on innovativeness of an agent as measured by time-of-adoption is classified into thus: (1) early adopters, (2) early majority, (3) late majority, and (4) laggards. With threshold values computed at personal and system-levels, it is imperative to for these classifications as: *early* adopters are agents whose time-of-adoption is greater than one standard deviation earlier than the average time-of-adoption. *Early* and *late* majorities are agents whose time-of-adoption is bounded by a standard deviation earlier/later than the average. Finally, *laggards* are agents who adopted later than one standard deviation from the mean (Ryan and Gross, 1950; Beal and Bohlen, 1955; Rogers, 1983).

Personal network threshold adopter categories may be created by partitioning the network threshold distribution in the same manner described for time-of-adoption adopter categories. Specifically, very low network threshold value individuals have personal network thresholds one standard deviation lower than the average threshold. Low and high network threshold individuals have personal network thresholds bounded by one standard deviation less than and greater than average. Finally, very high network threshold individuals have personal network thresholds one standard deviation greater than average. The average threshold is the mean threshold value for the community or graph system or the network (as case may be). Adopter categories provide a mechanism for audience segmentation, the comparison of research results, and the summation of research findings (Rogers, 1983). Also, these adopter categories were created to compare early adopters with later adopters to determine differences in their social and personal characteristics, communication behavior, and opinion leadership. One of the primary research findings of diffusion research was that early adopters had more sources of external influence.

4. Proposed Stochastic Graph-Based Model

The study proposes the stochastic graph model to test role of external influences on the adoption of innovation as clarified via threshold values that measures and estimates the expected number of final adopters in time. Also, this dual classification permits specifications of how external and/or interpersonal influences flows through the system to govern the diffusion process.

The nature of the problem determines the graph type used such as scale-free or small world graphs, and test the extent of local, nodal exogenous and dependency endogenous feats as contained therein. These helps to explain observed global structures on a graph such as path dependence, edgewise shared partnership ties, cluster coefficient etc – as captured via nodal attributes, its patterns of shared partners and k-triangles that are relatively local structures (Frank and Strauss, 1986; Wasserman and Pattison, 1996; Robins et al, 2007; Snijders et al., 2006; Goodreau, 2007).

For the proposed framework, we adopt the Exponential Random Graph Model (ERGM) such that it holds true all the feats as noted above. Thus, if a random graph G with *n* nodes and *m* possible ties between nodes i and j is denoted by a random variable G_{ij}, these parameters will hold as thus: *Its Probability Distribution* in ERGM is defined thus:

$$P_{\theta,G}(G = g) = \frac{\exp[\theta^t u(g)]}{c(\theta,G)} \quad (8)$$

u(g) is vector statistics of realization G, C(θ ,G) is a normalization function to ensure distribution sums to 1, and θ

is vector parameters of particular ERGM, found via Maximum likelihood in Markov Chain sampling to helps to draw realizations of G via distribution $P_{\theta,G}$ (Robins et al, 2007).

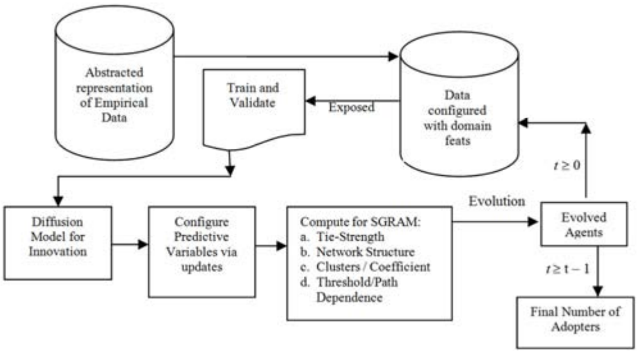

Fig. 1. *Experimental Evolution for Stochastic Social Graph Model (SGRAM).*

Even with tie-strengths as they build into personal networks, actors vary in their willingness to take risks in adopting new idea, product or behaviour. Thus, most actors are reluctant to do same as they will rather wait until others have tried the idea, behaviour or product first (Valente, 1996) – this represents the quantity *threshold* and consists of 4-major feats: (i) actor $i \in M$ with $P_i(G) = \{j \mid (j,i) \in \varepsilon\}$ to indicate its neighbours that can influence i to adopt the said innovation, (b) $\Phi(l)U_{l=0}^{t-1}$ is the rate at which an actor i has neighbours who are also members of the *seed*-set node that enables actor i to engage in the consideration that adoption is worth it or not, (c) $|N_i(G)|$ - is the entire network of actors who can influence actor i to adopt the innovation, and (d) set $A(l)$as number of actors that will adopt the innovation. The set $\Phi([0,1])$ have agents yet to be exposed at a given $t = 0$; But, who will be persuaded by their neighbours that adoption is worth considering – after which they will adopt the innovation as in Eq. 5. The algorithm is as thus:

Algorithm Local Emergent Feats {n, m, M, P}

1. Initialize number of nodes n; number of Ties m
2. Set Initial Tie Strength = 10+, Cluster Structure = 25+;
3. Set Network Structure as function of clusters
4. Initialize Graph with agents via: $P_{\theta,G}(G = g) = \frac{\exp[\theta^t u(g)]}{c(\theta,G)}$
5. Set Agent position with Max and Min bounds of expected number of final adopters as a function of Motion M.
6. Randomly select nodes for seedset choice
7. While Node are yet to be Exposed
8. Choose agents position in seedset as best position in graph
9. Initialize current Agent position = {Mmin + rand(Pmax – Pmin)}
10. Compute Threshold, Path dependence for all nodes in G
11. Set Pairwise Variables Interactions (+Di)
12. Set Predictive Variables for task (+Ri)
13. Compute Agent Tie Strength in{+Ri, +Di, +NiG, +Eli)
14. For Each Agent, Node or Actor i: Do {
15. If seedset is member of an agent's personal network
16. Then node.list.append(seedset node)
17. End If
18. End For Each
19. Compute Network Structure given Ni as:

$$N_i(G) = P_{\theta,G} + \lambda_0\mu_L + \lambda_1 Med_L + \sum_{t=0}^{t-1}\sum_{i\in L}\lambda_t(s-\mu_L)^t + \lambda_5 Min_L + \lambda_6 Max_L$$

1. Compute Agent's new disposition from exposure at t ≥ 0:

$$M_{new} = w^*M_{old} + c_1^* \text{ rand() } * [(P_i)/T] + c_2 * \text{ rand() } * [(P_n)/T)]$$

2. Updating agents' positions as $P_{new} = (P_{old} - M_{new})$
3. //continue till all nodes are exposed, then stop.
4. Stop if stop criterion is met

As the diffusion process continues at t ≥ 0, agents are first exposed (with the existence of seed-set) to allow them form their perception and behaviour about the innovation. In time, as more of the agents are exposed to innovation, they make preferences – and based on their threshold, adopt the new innovation as well as form clusters and cliques via learning outcome. These strengthen tie-relationship between agents and help them better retain information within their memory in time as the system continues in its search for optimality. The random exchange in an agent's personal network allows knowledge swap – so as to yield in time agents with a new set of disposition (consider adoption). As more adoption is encountered, more agents continue to learn/retain contents within their memories that better their personal network via community-influences.

Each exposure yields an updated number of final adopters as its optimal solution in time via recomputed threshold value for exposed actors. We note that: (a) the agent position range is normalized between [0-1] dividing it by maximum range of agents, (b) each position randomly determines swap type needed for adoption rate, and (c) positions are reset and these recomputed new values will eventually reflect system threshold. With each solution found, model restarts with another randomly selected point for the planted seed-set choice in the graph space.

Agents with threshold value above 0.5 are chosen. Process continues till all agents are exposed time ≥ t – 1 at which all agents will have a threshold of 1 for the diffusion process or the nodes are continuously re-evaluated till an agent is found of threshold lesser than or equal to start-off threshold value (these form the stopping criterion for the model). At which point the solution is reached.

5. Result Findings and Discussion

Two major external sources that were of great influence on adoption for the diffusion process are: (a) cosmopolitan actions, and (b) communication media. Cosmopolitan actions and media consumption provide individuals with earlier awareness of an innovation (Becker, 1970; Fischer, 1978) as well as freedom from system norms (Menzel, 1960) so that they can become earlier adopters and proponents of an innovation. In communities such as art and science, the norm is for innovative behavior and so external influence may operate differently (Michaelson, 1993). Rogers (1993) classification includes innovators as agents who adopt extremely early or are the very first to adopt. In this project however, they are included with early adopters since they represent a small fraction of the sample (10.4%). A cosmopolitan agent is an individual who is oriented to the world outside of his/her local social system (Merton, 1968) and relates his/her local social system to the larger social environment by providing his personal network and entire social system with links to outside data (Weimann. 1982; Gouldner, 1958; Davis, 1961).

5.1. Model Performance Evaluation

Performance is evaluated via mean square error, mean absolute error, mean relative error, coefficient of efficiency and coefficient of determination (R^2). MSE, MRE, MAE have an ideal value 0; while COE and R^2 have ideal value 1 as in Eq. 9 – Eq. 13. Y_{pi}/Y_{oi} are *predicted* and *observed* outputs with n observations as:

$$MSE = {}^{1}/_{n} \sum_{i=1}^{m} \{(Y_{pre} - Y_{obs})^2\}^{1/2} \tag{9}$$

$$MAE = {}^{1}/_{n} \sum_{i=1}^{m} |Y_{pre} - Y_{obs}| \tag{10}$$

$$MRE = {}^{1}/_{n} \sum_{i=1}^{m} \frac{|Y_{pre} - Y_{obs}|}{Y_{obs}} \tag{11}$$

$$COE = 1 - \frac{(Y_{obs} - Y_{pre})^2}{(Y_{pre} - Y_{obs})^2} \tag{12}$$

$$R^2 = \frac{([Y_{obs} - (1 - Y_{obs})][Y_{pre} - (1 - Y_{pre}])^2}{([Y_{obs} - (1 - Y_{obs})])^2([Y_{pre} - (1 - Y_{pre}])^2} \tag{13}$$

Table 2. *Model Performance of Parameters.*

Strategy	MAE	MRE	MSE	COE	R^2
$+R_i$	0.021	0.102	0.023	0.781	0.766
$+D_i$	0.012	0.110	0.036	0.753	0.821
$+N_iG$	0.01	0.192	0.029	0.688	0.812
$+S_i$	0.10	0.110	0.032	0.871	0.901

Validation is more of a scientific discussion that ambiguous results as improperly applied, often impedes. Our study aims to minimize confusion in social graph diffusion using agent based modeling as we aim to establish parameters for its measurement (Ojugo et al, 2012). Model's performance aim to exploits predictive variables in network structure, tie strength and pairwise interactions – as the model's achieves its coefficient of determination R^2 with MSE, MRE and MAE at $p < 0.1$ on a continuous 0–1 scale, where 0 is weakest and 1 is strongest. On the average, R^2 for the parameters as predicted shows seventh-tenth of its true-value. Its error interval tightens at end of the continuum to suggest a strong evidence of interaction between all these dimensions at $p < 0.1$. Parameters as structural dimension, continues to play a minor role (in its linear form factor) but has an important modulating role to imply that relationships matter; Thus, are filtered through clusters before impacting on tie-strength, network structure and eventually, lead to the consequent adoption/rejection of the innovation by agents as predicted to yield the expected number of final adopters.

5.2. Result Presentation and Discussion

Table 3. *Tie strength on Time of adoption.*

	Dependent Variables	μ	+Ri	+Ni	+Di	ELi
1	When was your first or last communication? How long have you known?(Duration)	0.67	0.94	0.87	0.89	0.21
2	How strong is your relationship? (Intensity)	0.47	0.87	0.78	0.89	0.10
3	On Education, political and occupational differences: how helpful has project been? (Social Distance)	0.23	0.42	0.34	0.43	0.19
4	How comfortable are you with, to loan from him/her? (Emotional Support)	0.43	0.90	0.92	0.95	0.21
5	Recency in communication, relationship status and how would you feel if unfriended by him/her? (Intimacy)	0.28	0.81	0.65	0.78	0.10
6	How important is program to you and for you to bring your friends? (Reciprocity)	0.38	0.81	0.76	0.72	0.13
7	How many mutual friends do you share, do you share the same interest (interest overlap) and how many clusters do you both belong to in-common? (Structural Distance)	0.31	0.80	0.82	0.86	0.34

Many studies with various manifestations of tie strength (Krackhardt, 1990; Haythornthwaite, 2002), allows capture of multiplicity and diversity views of participants who were asked to answer 7-tie strength questions. Participants moved a slider along a continuum to rate a friend as illustrated in table 3. The continuum was chosen rather than a discrete scale for three reasons: (a) Granovetter conjectured that tie-strength may and is in fact continuous (Granovetter, 1973). This study and many others are not poised to resolve if tie-strength is discrete of not, and have also not specified how many discrete tie-strength levels exist. The continuum to this end, bypasses that problem, (b) a continuum lends itself to standard modeling techniques, which is both stochastic and evolutionary, and (c) lastly, its application can round a continuous model's predictions to discrete-levels when and as needed and appropriate (Gilbert and Karahalois, 2009).

The adoption of a linear combiner model for tie strength prediction allows us to take advantage of the full dataset and explain the results once built. s_i is tie-strength of agents i, R_i is number of predictive variables for task at hand, e_i is the error term in tie strength, $N_i(G)$ is network structure. *Di* all pairwise relations between the 21-variables as included in predictive model with 90% or greater completion rate. This is further resolved via reciprocity. There are 3-major variables for the 7-dimensions (for this case) and we use a linear combiner to force more variables than required data points into the model with a 90% system-threshold to ensure all dimensions are adequately represented to explore interactions between the dimensions of tie strength.

The results shows that where *seedset* of agents that can influence others exists, and are uniformly distributed over the graph, the study explores the relationship of the expected number of final adopter (our metrics), clustering effects, threshold values with path-dependence, network structure and seed cardinality – to suggest that the highly clustered networks with more seed-sets allowed for easier diffusion of innovation. While clustering and high clustering coefficients promotes diffusion where there exists seed node inside an agent's personal networks and social system, such cohesive and highly clustered sets are also quite difficult to penetrate when they are not targeted during initial seeding phase. In some cases, sampling strategy used to generate the agents for the diffusion process can create some errors in the time of adoption, alongside the external influences that acts on the system etc. These external influences such as finance, tribalism, race and gender etc – can also act as a motivator during each transaction that involves corruption ethic, especially with such actors not being monitored. This will alter the outcome of the diffusion process significantly. A major reason why project was kept at 18months is to seek time convergence and the acceptability of the initiative – that will help change actor perception, disposition and behaviour and consequently allow complete diffusion as actors create stronger ties. External influence and path dependence only accounts for about 11.8% of deflection from the true purpose (as in table 3).

Table 4. *Convergence of Expected number of final Adopters with External Influence on Adoption.*

Adopter Category	Ext. Infl.	Personal network: direct ties					
		μ	δ	Low	Ave	High	System Total
Early	1.1	0.76	0.31	1.9	2.7	9.6	14.2%
Early Majority	2.1	0.57	0.23	4.8	5.2	10.6	20.6%
Late Majority	2.1	0.46	0.21	8.7	6.1	9.0	23.8%
Laggards	6.5	0.32	0.20	11.9	16.1	13.4	41.4%
Personal Network Total	11.8			27.3	30.1	42.6	100%

Table 4 shows percentages for adopter categories relative to both the entire system and to an agent's personal networks at 90% significance. Resultant variables are associated with each other. The system notes that one's time-of-adoption is associated with proportion of adopters in the social system – and thus, associated with proportion in an agent's personal network. 60% of youths were classified identically in both personal and system thresholds – to represent proportion of actors who are more innovative relative to the system than to their network or more innovative relative to their personal network than to the social system. Also, 20.8% of youths are more innovative relative to the social system since they adopted in the early adopter phase, yet waited until some portion of their personal network adopted. All the youths in *laggard* phase have either very low or very high thresholds (Valente, 1996).

Actors that fall within the laggard resulting to about 40% are not innovation either to their personal network or to the entire social system. Laggards represent a skewness from a considerable proportion of non-adopters in the respective datasets and accounts for mainly about 22.4% of the youths. A reason for this skewness is based on the projection that all agents will eventually adopt given a certain time that exceeds the stop criterion of time t = t – 1. At which time, the innovations will be completely adopted. It is also noted that with some other forms of diffusion of innovation, some adopter agents remain indifferent and thus, classified as *laggards*.

Also, the results shows that the external influence scores for each of the categories (low, average and high) adopters vary for agents who are innovative relative to the system. External influences here ranged from campaign, tribal/race sentiments, cosmopolitan effect/action, finance, gender etc. Youths who are most innovative relative to the social system and have very low thresholds based on campaign exposure. Results also indicate that youths who were early adopters relative to the system and had low thresholds subscribed to an average of 18.2%. Analysis of variance conducted to test the association between innovativeness variables and degree of external influence was consistent across training datasets, indicating that external influence operated the same way in the study. Youths campaign exposure score were associated to innovativeness dimensions and interaction term.

5.3. Theoretical Implications

With graph-based model, there are still a lot more variance to understand such as predictive variables and "behind-the-scenes" data. Though, the addition of more data to the task at hand may not solve it – but, we have aimed to define some parameters that helps predict social graph. There is the need to develop and standardize novel indicator parameters for social-graph models, which in turn raises new questions for the theory. In modeling tie strength in social graph, what important feats and parameters are necessary predictors and used in prediction of threshold? What upper/lower limits are to be set for tie strength predictability?

We believe our work makes some important contributions to the theory as thus: (a) extending tie-strength dimension as manifested in all social-graph models, (b) defining network structure dimension as a function of probability distribution of agents in problem space, (c) all dimensions modeled as a continuous value, (d) our result extends the realization of how structural dimension in predictive variables used in task help modulates other dimensions within, by filtering agent relationships through cliques and clusters, and (e) previous works assumed either the presence or absence of a link in the graph without recourse to the properties of the link itself. Thus, introducing a complete tie strength model into social graph or network with even real-world data may yield a novel conclusion (Gilbert and Karahalois, 2009).

5.4. Rationale for Model

Many studies focuses on dyadic interaction and relation between agents so as to measure parameters such as number of agents *m* and ties *n* that amounts to other parameter such as network structure, threshold, path-dependence, clusters and communities, cohesion, external influences etc. These as explored by the model as agents' local emergent feats – are then encoded within the parameter tie-strength to help account for final number of adopter in diffusion process cum propagation analysis. It help determine an agent's position, behaviour and preference towards an innovation as well as aid proper decision making in graph figuration cum settings via agent-based model.

This model exploits interactions amongst an agent's local emergent feats to orchestrate a relation that predicts strong tie-strength based on the agent's threshold, path-dependency and external influence effects (from a highly stylised form) in order to capture global behavioural pattern as enactment process that aid retrospective decision making. Path-dependent and external influence are effects on agent that occurs in response to events that may disrupt diffusion – just as tie-strength and network structures prediction helps ascertain on a social network, the agent's preference as it determines the actual behaviour emerging in such an agent overtime. Thus, network studies must focus on dyadic status and relationships as well as establish parameters through training and validation with cross-sectional data.

6. Conclusion and Recommendations

Models are predictive, educational tools to aid experts and researchers compile existing knowledge about a task. They serve as vehicle to communicate hypotheses, a means to investigate parameters crucial in estimation as well as help us better understand a problem domain. Simple models may not provide enough new data, whereas complex models may not be understood. Model implementation as an intellectual tool, requires less accurate numeric agreement in predicted versus observed values as it displays feat of interest with its probabilities. But, rather requires feedback mechanism as more important. Only models that are understandable and easily manageable are fully explored. Thus, modelers must balance complexity and simplicity, which is crucial to studying the relevant processes of how a model works.

Acemoglu et al (2012) also completely characterized a set of final adopters as a function of cohesion, clustering, path dependence, threshold, seed set and other feats (uniformly distributed over the social graph). Results show that a highly clustered network is more advantageous over less structured graphs with large numbers of random links. Clusters and cohesive cliques promote diffusions faster with existence of seed node inside it. However, they are harder to penetrate if they are not targeted during initial seeding phase.

References

[1] Abraham, A., (2005). *Handbook of Measuring System Design*, John Wiley and Sons Ltd, ISBN: 0-470-02143-8, pp. 901 – 918.

[2] Alpaydin, E., (2010). *Introduction to Machine Learning*, McGraw Hill publications, ISBN: 0070428077, NJ

[3] Axelrod R., (1997). *The Complexity of Cooperation*. Princeton, NJ: Princeton Univ. Press

[4] Beal, G.M and Bolen, J.M., (1955). *How farm people accept new ideas*, Ames, IA: Cooperative Extension Service Report 15.

[5] Becker, M.H., (1970). *Sociometric location and innovativeness: Reformulation and extension of the diffusion model*, American Sociological Review, 35, p267-282.

[6] Burt, R.S., (2004). *Structural Holes and Good Ideas*, American Journal of Sociology, 110(2), p349–399.

[7] Clauset, A., Shalizi, C.R and Newman, M.E.J., (2009). *Power-law distributions in empirical data*, *Siam Review,* 51(4), p661–703.

[8] Coleman J.S., (1990). *Foundations of Social Theory*. Cambridge, MA: Harvard Univ. Press.

[9] David, P.C., (2007). *Path dependence – a foundational concept for historical social sciences*, Climetrica – Journal of Historical Economics and Econometric History, 1(2).

[10] Davis, J.A., (1961). *Locals and cosmopolitans in effects in American graduate schools*, Int. Journal of Comparative Sociology, 2(2), p212-223.

[11] Dozier, D.M., (1977). *Communication Networks and threshold role in the adoption of innovations*, PhD Thesis, Stanford University.

[12] Durkheim, E., (1982). *The Rules of the Sociological Method*, London: Macmillan.

[13] Epstein, J and Axtell R., (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Cambridge, MA: MIT Press.

[14] Fischer, C.S., (1978). *Urban-to-rural diffusion of opinions in contemporary America*, American J. of Sociology, 84, p151-159.

[15] Friedkin, N. E., (1980). *A Test of Structural Features of Granovetter's Strength of Weak Ties Theory*, Social Networks, 2, p411–422.

[16] Gilbert, E., Karahalios, K and Dresden, Y., (2008). *The Network in the Garden: An Empirical Analysis of Social Media in Rural Life*, Proceedings of CHI, p1603–1612.

[17] Gilbert, E and Karahalois, K., (2009). *Predicting tie strengths with social media*, J. Computer and Human Interface, 15, p76-97, ACM 978-1-60558-246-7/09/04.

[18] Golbeck, J. (2013). *Analyzing the Social Web*, Morgan Kaufmann, ISBN: 0-12-405856-6.

[19] Gouldner, A., (1958). *Cosmopolitan5 and locals: Toward an analysis of latent social roles*, Administrative Science Quarterly 1 and 2, p281.

[20] Granovetter, M., (1973). *The Strength of Weak Ties*, The American Journal of Sociology, *78*(6), p1360–1380.

[21] Granovetter, M. (1978). *Threshold Models of Collective Behavior*. American J. Sociology, 83(6), p1420–1443. doi:10.1086/226707, (http://dx.doi.org/10.1086%2F226707). JSTOR 2778111

[22] Granovetter, M. (1983). *Strength of Weak Ties: A Network Theory Revisited*, Sociological Theory, 1, p201–233, doi:10.2307/202051, (//www.jstor.org/stable/202051).

[23] Granovetter, M. (1985). *Economic action and social structure: The problem of embeddedness,* American J. Sociology, 91(3), p481–510, doi:10.1086/228311.

[24] Handcock, M.S and Gile, K.J., (2009). *Modeling social networks from sampled data*,* Annals of Applied Statistics, arXiv: math.PR/00000.

[25] Haythornthwaite, C., (2002). *Strong, Weak, and Latent Ties and the Impact of New Media,* Information Society, 18(5), p385–401.

[26] Haythornthwaite, C and Wellman, B., (1998). *Work, Friendship, and Media Use for Information Exchange in a Networked Organization*, Journal of American Sociology and Information Science, 49(12), p1101–1114.

[27] Krackhardt, D. (1990). *The Strength of Strong Ties: Importance of Philos*, In N. Nohria and R. Eccles., *Networks and Organizations: Structure, Form and Action* (p216–239), Harvard Biz School Press.

[28] Krackhardt, D and Stern, R.N., (1988). *Informal Networks and Organizational Crises: Experimental Simulation*, Social Psychology Quarterly, 51(2), p123–140.

[29] Kaufman S., (1996). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. UK: Oxford Univ. Press.

[30] Lin, N., Ensel, W.M and Dayton, P.W., (1981). *Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment*, American Sociological Review, *46*(4), p393–405.

[31] Macy M and Willer, J., (2002). *From factor to actors: computational sociology and agent based model*, Annual Review Sociology, 28, p143–166, doi: 10.1146/annurev.soc.28.110601.141117.

[32] Marin, A., (1981). *Are Respondents More Likely To List Alters with Certain Characteristics?: Implications for Name Generator Data*, Social Networks, *26*(4), 289–307.

[33] Marsden, P. V and Campbell, K.E., (1990). *Measuring Tie Strength*, Social Forces, 63(2), p482–501.

[34] Menzel, H., (1960). *Innovation, integration and marginality: a survey of physicians*, American Sociological Review, 25, p704-713.

[35] Merton, R.K., (1968). *Patterns of influence: Local and cosmopolitan influentials*. Reprinted from: P.F. Lazarsfeld and F.N. Stanton (Ed.) 1948-1949. *Communication Research*. NY: Harper and Brothers.

[36] Michaelson, A.G., (l993). *The development of a scientific specialty as diffusion through social relations: The case of role analysis*, Social Networks, 15, p217-236.

[37] Mitchell, T.M., (1997). *Machine Learning*, McGraw Hill publications, ISBN: 0070428077, New Jersey.

[38] Newman, M.E.J., (2003a). *Mixing patterns in networks*. Physical Review E, 67-026126, p90-102.

[39] Newman, M.E.J, (2003b). *The structure and function of complex networks*. SIAM Reviews, 45(2), p167-179.

[40] Ojugo, A.A., A.O. Eboka., E.O. Okonta., E.R. Yoro and F.O. Aghware., (2012). *Genetic algorithm intrusion detection system*, Journal of Emerging Trends in Computing and Information Systems, 3(8): 1182-1194

[41] Ojugo, A.A., (2013). *Virus propagation on time varying graphs,* Technical-Report, Centre for High Performance and Dynamic Computing, TRON-03-2013-01, Federal University of Petroleum Resources, Nigeria, p24-37.

[42] Ojugo, A.A., (2014). *Malware propagation on time varying network: comparative study of machine learning techniques and frameworks,* Int. J. Modern Education and Computer Science, 8: 25-33, doi: 10.5815/ijmecs.2014.08.04.

[43] Reynolds C.W., (1987). *Flocks, herds, and schools: a distributed behavioral model,* Computer Graphics, 21, p25–34

[44] Rogers, E.M., (1983). *Diffusion of innovation*, 3rd Ed. New York: Free Press.

[45] Rogers, E.M and Kincaid, D.L., (1993). *Communication Networks: New Paradigm for Research*, NY: Free Press.

[46] Sala, A., Cao, N., Wilson, C., Zablit, R., Zheng, H and Zhao, B.Y., (2010). *Measurement calibrated graph models for social network experiments*, IW3C2, ACM 978-1-60558-799-8/10/04.

[47] Schnettler, S., (2009). *A small world on feet of clay? A comparison of empirical small-world studies against best-practice criteria*, Social Networks, 31(3), p179-189, doi:10.1016/j.socnet.2008.12.005.

[48] Scott, J.P., (2000). *Social network analysis: A Handbook* (2nd Ed). Thousand Oaks, CA: Sage Publications.

[49] Simon H. 1998. *The Sciences of the Artificial*. Cambridge, MA: MIT Press

[50] Smith T.S and Stevens, G.T., (1999). *The architecture of small networks: strong interaction and dynamic organization in small social systems*, American Sociological Review, 64, p403–20

[51] Strang, D and Macy, M., (2001). *In Search of Excellence:" fads, success stories, and adaptive emulation,* American Sociological Review, 107(1), p147.

[52] Toivonen, R., Kovanena, L., Kiveläa, M., Onnela, J.K., Saramäkia, J and Kaskia, K., (2009). *A comparative study of social networks models: network evolution and nodal attributes models*, Social Networks, 31, p240-254, doi:10.1016/j.socnet.2009.06.004.

[53] Valente, T.W., (1996). *Social network thresholds in the diffusion of innovation*, Social Networks, 18, p69-89, SSDI 0378-8733(95)00256

[54] Wasserman, S and Faust, K., (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press, p1–27, ISBN 97805213870-71.

[55] Weimann, G., (1982). *On the importance of marginality: one more step into the two-step flow of communication*, American Sociological Review, 47, p764-773.

[56] Wellman, B and Wortley, S., (1990). *Different Strokes from Different Folks: Community Ties and Social Support*, American Journal of Sociology, 96(3), p558–588.

[57] Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P and Zhao, B.Y., (2009). *User interactions in social networks and their implications*. In *Proc. of EuroSys* (April 2009).