EcoSupply: A Machine Learning Framework for Analyzing the Impact of Ecosystem on Global Supply Chain Dynamics

Vikas K. Garg[1] and N. Viswanadham[2]

[1] IBM Research - India
vikaskga@in.ibm.com
[2] Global Logistics and Manufacturing Strategies (GLAMS)
Indian School of Business (ISB), Hyderabad, India
N_Viswanadham@isb.edu

Abstract. A global supply chain spans several regions and countries across the globe. A tremendous spurt in the extent of globalization has necessitated the need for modeling global supply chains in place of the conventional supply chains. In this paper, we propose a framework, *EcoSupply*, to analyze the supply chain ecosystem in a probabilistic setting unlike the existing methodologies, which presume a deterministic context. EcoSupply keeps track of the previous observations in order to facilitate improved prediction about the influence of uncertainties in the ecosystem, and provides a coherent mathematical exposition to construe the new associations, among the different supply chain stakeholders, in place of the existing links. To the best of our knowledge, EcoSupply is the first machine learning based paradigm to incorporate the dynamics of global supply chains.

Keywords: Supply Chains, Global Sourcing, Machine Learning.

1 Introduction

Global outsourcing has acquired a central role in the contemporary manufacturing and service industries. Multinational companies invest at different locations across the globe to gain competitive advantage by exploring new markets, availing cutting edge technology, and harnessing skills at sustainable costs. Therefore, the need for making an effective decision regarding the selection of locations and global business partners from a plausible set of candidates can not be overemphasized.

The literature abounds in techniques for modeling the supply chain formation. Walsh et. al [1] proposed a combinatorial protocol, consisting of a one-shot auction and a strategic bidding policy, to study the negotiations on production relationships among multiple levels of production in a distributed setting. Prior to this work, auction mechanisms were proposed, such as in [2], to address the complementarities or the mutual dependencies among values of obtaining inputs and producing outputs. Typically, a global supply chain is the result of trade by

K. Deb et al. (Eds.): SEAL 2010, LNCS 6457, pp. 677–686, 2010.

a firm across national borders by means of either foreign direct investment (FDI) or outsourcing, though other levels of operational strategies such as licensing, joint venture, and acquisition etc. also exist. Consequently, a lot of effort has gone into investigating the decision of firms to trade through FDI or outsourcing, see for instance, [3], [4]. Most of these models can not be quantitatively analyzed and proffer only a high level insight into the decision making process.

Tax has a significant impact on SCF, as the product material moves across boundaries. Certain parts of the world offer special economic zones - also known as free trade zones - where goods bound for export can be manufactured, assembled, and stored with attractive tax holidays. Therefore, substantial research has gone into integrating taxes and other regulatory factors in the global supply chain design ([5], [6]). Recently, a mixed integer non-linear programming model that incorporates the import and export tax liabilities at various stages of the global supply chain has been proposed [7]. Besides tax, there are certain other factors with positive (for example, acquaintance) or negative (e.g. economic and cultural heterogeneity) influence that have a marked influence on the overall supply chain formation. However, these factors have been overlooked thus far in the literature.

1.1 Motivation

The literature abounds in expository research on supply chain formation (SCF) and network planning. However, almost all of these techniques analyze the problem of selecting an alternative at a given stage using a deterministic cost model while neglecting altogether the uncertainty in the surrounding ecosystem, which encompasses all the factors that might influence the supply chain formation. For instance, there are certain factors in most supply chain ecosystems, such as infrastructure, local demand and proximity to key markets, availability of skilled labor, inventory handling facilities, government regulations and incentives, financial costs (e.g. in acquisition of land), transportation, and tax and freight considerations, etc. While some of these factors, notably tax considerations and inventory handling costs, have been incorporated in the existing models, a vast majority of these factors still remains unaccounted. Furthermore, most of the sub-factors that determine these factors may change over a period of time, thereby triggering a change in the impact of these factors. Therefore, we believe there is a need for a generic probabilistic framework that seamlessly incorporates and integrates these factors for understanding the dynamics of the supply chains. This adaptive modeling of supply chains is fundamental to explaining the replacement of an extant end-to-end supply chain with a new one, as the different factors governing the SCF change over time. In this work, we explicate this dynamic aspect of supply chains using a statistical model EcoSupply.

2 Problem Definition and Notation

Consider a multi-stage global supply chain network, where each stage represents an activity such as production or assembly. We assume that the Supply chain has

N stages: $S_1, S_2, \ldots, S_N$. There are k_i alternatives, $S_{i1}, S_{i2}, \ldots, S_{ik_i}$, at any stage S_i to accomplish the activity of that stage. Each alternative q at a stage p is expressed as a d-dimensional observation or feature vector of factors x_{pq}, with the data observation corresponding to a factor r denoted by x_{pqr}, whereby the probability of a factor r being favorable is given by $\theta_{pqr}, 0 \leq \theta_{pqr} \leq 1, r \in \{1, 2, \ldots, d\}$. Let D_{pq} represent a set of n d-dimensional data vectors corresponding to the q^{th} alternative in stage S_p, and D_{pqr} represent a set of n samples corresponding to factor r, assumed to be independent and identically distributed (i.i.d), $\{x^1_{pqr}, x^2_{pqr}, \ldots, x^n_{pqr}, x'_{pqr}\}$. Further, let τ_{pqr} denote the threshold above which a factor r is perceived favorable, and w_{pqr} and $C^{D_{pqr}}_{pqr}$ denote respectively the weight or perceived importance of r, and the estimated cost associated with r based on D_{pqr}, at the alternative q in the stage p. Finally, let C^{init}_{pq} and $C^{D_{pq}}_{pq}$ denote the initial unaccounted cost (which disregards the impact of factors), and the estimated cost based on D_{pq}, taking into consideration the ecosystem, if alternative q is chosen in stage p.

Then, the problem of probabilistic modeling of SCF is formulated as follows: find the probability of any supply chain, $SC = A_1, A_2, \ldots, A_N$, formed by choosing an alternative A_i from each stage $1 \leq i \leq N$. Intuitively, the greater this probability, the more likely the formation of SC, compared to any other supply chain. Furthermore, this probability might change over time, as more data is accumulated or the impact of various factors varies.

3 The EcoSupply Model

Two types of factors need to be considered: a) the factors local to an alternative, and b) the factors governed by a pair of alternatives at successive stages in the supply chain network. Below, we describe how these factors are modeled using the EcoSupply framework.

3.1 Modeling the Impact of Factors Specific to an Alternative

At any instant of time, each of the underlying factors in the ecosystem can be considered as being favorable or unfavorable towards selection of a particular alternative at a particular instant of time, e.g. there might be a fear of shortage in supply of raw materials at a particular alternative deeming a high cost for that alternative. Our aim is to continually learn the favorable probabilities as more data is accumulated over time. Each of the factors in the ecosystem can be perceived as Bernoulli variables representing unknown probability distributions. Then, the estimate for an observation x_{pqr} (which takes one of the two values: 1(favorable) or 0(unfavorable)) conditioned on the parameter θ_{pqr} is given by,

$$P(x_{pqr}|\theta_{pqr}) = \theta^{x_{pqr}}_{pqr}(1 - \theta_{pqr})^{1-x_{pqr}} \tag{1}$$

Then, we have the following result.

Lemma 1. *Let $D_{pqr} = \{x^1_{pqr}, x^2_{pqr}, \ldots, x^n_{pqr}\}$ be a set of n i.i.d samples drawn according to a probability distribution characterized by θ_{pqr}. If θ_{pqr} has a uniform prior distribution, then*

$$P(x_{pqr}|D_{pqr}) = \left(\frac{s^{D_{pqr}}_{pqr} + 1}{n+2}\right)^{x_{pqr}} \left(1 - \frac{s^{D_{pqr}}_{pqr} + 1}{n+2}\right)^{1-x_{pqr}}$$

where $s^{D_{pqr}}_{pqr} = \sum_{j=1}^{n} x^j_{pqr}$

Proof.

$$\begin{aligned} P(D_{pqr}|\theta_{pqr}) &= P(x^1_{pqr}, x^2_{pqr}, \ldots, x^n_{pqr}|\theta_{pqr}) \\ &= P(x^1_{pqr}|\theta_{pqr})P(x^2_{pqr}|\theta_{pqr}) \ldots P(x^n_{pqr}|\theta_{pqr}) \\ &= \theta_{pqr}^{\sum_{j=1}^{n} x^j_{pqr}} + (1-\theta_{pqr})^{\sum_{j=1}^{n}(1-x^j_{pqr})} \qquad [using\ (1)] \\ &= \theta_{pqr}^{s^{D_{pqr}}_{pqr}} (1-\theta_{pqr})^{n-s^{D_{pqr}}_{pqr}} \end{aligned} \tag{2}$$

Now,

$$p(\theta_{pqr}|D_{pqr}) = \frac{p(D_{pqr}|\theta_{pqr})p(\theta_{pqr})}{\int_{\theta_{pqr}} p(D_{pqr}|\theta_{pqr})p(\theta_{pqr})\, d\theta_{pqr}}$$

In the absence of any prior knowledge about θ_{pqr}, assuming a uniform distribution[1] in the interval [0, 1], we obtain,

$$\begin{aligned} p(\theta_{pqr}|D_{pqr}) &= \frac{p(D_{pqr}|\theta_{pqr})}{\int_0^1 p(D_{pqr}|\theta_{pqr})\, d\theta_{pqr}} \\ \Rightarrow P(x_{pqr}|D_{pqr}) &= \int_{\theta_{pqr}} P(x_{pqr}|\theta_{pqr})p(\theta_{pqr}|D_{pqr})\, d\theta_{pqr} \\ &= \int_0^1 P(x_{pqr}|\theta_{pqr}) \frac{p(D_{pqr}|\theta_{pqr})}{\int_0^1 p(D_{pqr}|\theta_{pqr})\, d\theta_{pqr}}\, d\theta_{pqr} \end{aligned} \tag{3}$$

Using (2),

$$\int_0^1 p(D_{pqr}|\theta_{pqr})\, d\theta_{pqr} = \int_0^1 \theta_{pqr}^{s^{D_{pqr}}_{pqr}} (1-\theta_{pqr})^{n-s^{D_{pqr}}_{pqr}}\, d\theta_{pqr}$$

[1] In general, each of the factors is dependent on several sub-factors, and may follow an arbitrary distribution, e.g. the supply of raw materials may not be uniform and may vary from time-to-time, depending on a change in the capability of the source or trade restrictions. This is not a very stringent assumption, for example, refer [9] for modeling a Gaussian prior on θ_{pqr}.

From the definition of beta function, for a, $b > 0$,

$$\beta(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

where $\Gamma(.)$ denotes the gamma function. Then, evaluating the gamma function on integral arguments, we get,

$$\int_0^1 p(D_{pqr}|\theta_{pqr})\, d\theta_{pqr} = \frac{\Gamma(s_{pqr}^{D_{pqr}}+1)\Gamma(n-s_{pqr}^{D_{pqr}}+1)}{\Gamma(n+2)} = \frac{s_{pqr}^{D_{pqr}}!(n-s_{pqr}^{D_{pqr}})!}{(n+1)!}$$

which in the light of (3) yields,

$$P(x_{pqr}|D_{pqr}) = \frac{(n+1)!}{s_{pqr}^{D_{pqr}}!(n-s_{pqr}^{D_{pqr}})!}\int_0^1 P(x_{pqr}|\theta_{pqr})p(D_{pqr}|\theta_{pqr})\, d\theta_{pqr}$$

$$= \frac{(n+1)!}{s_{pqr}^{D_{pqr}}!(n-s_{pqr}^{D_{pqr}})!}\int_0^1 \theta_{pqr}^{s_{pqr}^{D_{pqr}}+x_{pqr}}(1-\theta_{pqr})^{n-s_{pqr}^{D_{pqr}}+1-x_{pqr}}\, d\theta_{pqr} \;\; [using\ (1)\ and\ (2)]$$

$$= \frac{(s_{pqr}^{D_{pqr}}+x_{pqr})!(n-s_{pqr}^{D_{pqr}}+1-x_{pqr})!}{s_{pqr}^{D_{pqr}}!(n-s_{pqr}^{D_{pqr}})!(n+2)}$$

$$\Rightarrow P(x_{pqr}=1|D_{pqr}) = \frac{s_{pqr}^{D_{pqr}}+1}{n+2},$$

and,

$$P(x_{pqr}=0|D_{pqr}) = 1 - \frac{s_{pqr}^{D_{pqr}}+1}{n+2}$$

$$\Rightarrow P(x_{pqr}|D_{pqr}) = \left(\frac{s_{pqr}^{D_{pqr}}+1}{n+2}\right)^{x_{pqr}}\left(1-\frac{s_{pqr}^{D_{pqr}}+1}{n+2}\right)^{1-x_{pqr}}$$

In the next lemma, we show how the conditional density estimate can be incrementally updated on arrival of a new observation.

Lemma 2. *Let a new observation, x'_{pqr}, is recorded that results in an enhanced data set, $D'_{pqr} = D_{pqr}\bigcup\{x'_{pqr}\}$. Then, assuming the mutual independence of the d factors, the ratio of conditional probabilities,*

$$\frac{P(x_{pq}|D'_{pq})}{P(x_{pq}|D_{pq})} = \prod_{r=1}^{d}\left(\frac{n+2}{n+3}\right)\left[\frac{s_{pqr}^{D_{pqr}}+x'_{pqr}+1}{s_{pqr}^{D_{pqr}}+1}\right]^{x_{pqr}}\left[\frac{n-s_{pqr}^{D_{pqr}}+2}{n-s_{pqr}^{D_{pqr}}+1}\right]^{1-x_{pqr}}$$

Proof. It follows from Lemma 1 that

$$P(x_{pqr}|D'_{pqr}) = \left(\frac{s_{pqr}^{D'_{pqr}}+1}{n+3}\right)^{x_{pqr}}\left(1-\frac{s_{pqr}^{D'_{pqr}}+1}{n+3}\right)^{1-x_{pqr}}$$

$$= \left(\frac{s_{pqr}^{D_{pqr}} + x'_{pqr} + 1}{n+3}\right)^{x_{pqr}} \left(1 - \frac{s_{pqr}^{D_{pqr}} + x'_{pqr} + 1}{n+3}\right)^{1-x_{pqr}}$$

$$\Rightarrow \frac{P(x_{pqr}|D'_{pqr})}{P(x_{pqr}|D_{pqr})} = \left(\frac{n+2}{n+3}\right) \left[\frac{s_{pqr}^{D_{pqr}} + x'_{pqr} + 1}{s_{pqr}^{D_{pqr}} + 1}\right]^{x_{pqr}} \left[\frac{n - s_{pqr}^{D_{pqr}} + 2}{n - s_{pqr}^{D_{pqr}} + 1}\right]^{1-x_{pqr}}$$

Generalizing to the d-dimensional multivariate case by assuming that these d factors are mutually independent, we obtain,

$$P(x_{pq}|D'_{pq}) = P(x_{pq}|D_{pq}) \prod_{r=1}^{d} \left(\frac{n+2}{n+3}\right) \left[\frac{s_{pqr}^{D_{pqr}} + x'_{pqr} + 1}{s_{pqr}^{D_{pqr}} + 1}\right]^{x_{pqr}} \left[\frac{n - s_{pqr}^{D_{pqr}} + 2}{n - s_{pqr}^{D_{pqr}} + 1}\right]^{1-x_{pqr}} \tag{4}$$

Therefore, using Lemma 2, we can incrementally update the conditional density estimate on arrival of x'_{pqr}. Let τ_{pqr} be the threshold that determines if the factor r is favorable at alternative q in stage p. Then, one of the ways to compute the effective cost is given by,

$$C_{pq}^{D_{pq}} = C_{pq}^{init} \left[\sum_{j \in J_{pq}^{D_{pq}}} w_{pqj} e^{\frac{\tau_{pqj} - P(x_{pqj}=1|D_{pqj})}{m_{pqj}}} - |J_{pq}^{D_{pq}}| + 1 \right], \; where \tag{5}$$

$$J_{pq}^{D_{pq}} = \{j : \tau_{pqj} > P(x_{pqj} = 1|D_{pqj})\}$$

The weights w signify the importance of the different factors; the values τ can be adjusted to reflect the penalty in case of factors not meeting the desired threshold levels; and the scaling parameters m control the non-linearity of the model. Note that if a factor is deemed favorable with respect to the corresponding threshold, given the available data, then it does not add to the initial cost estimate, which disregards the ecosystem.

Theorem 1. *The overall estimated cost taking into account all the factors in the ecosystem at an alternative q in stage p based on D'_{pq}, for the cost model proposed in (5), is given by,*

$$C_{pq}^{D'_{pq}} = C_{pq}^{init} [\sum_{r \in J_{pq}^{D'_{pq}}} \left(I_{\tau_{pqr}}^{D_{pqr}} C_{pqr}^{D_{pqr}} + (1 - I_{\tau_{pqr}}^{D_{pqr}}) w_{pqr} \right)$$

$$e^{\frac{\tau_{pqr} - P(x_{pqr}=1|D_{pqr})}{m_{pqr}} \left\{ \frac{x'_{pqr}}{P(x_{pqr}=1|D_{pqr})} - I_{\tau_{pqr}}^{D_{pqr}} \right\}} - |J_{pq}^{D'_{pq}}| + 1] \tag{6}$$

Proof. Two cases are possible:

Case 1: $\tau_{pqr} > P(x_{pqr} = 1|D_{pqr})$ and $\tau_{pqr} > P(x_{pqruwc} = 1|D'_{pqr})$
Then, using

$$\frac{P(x_{pqr} = 1|D'_{pqr})}{P(x_{pqr} = 1|D_{pqr})} = \frac{(n+2)(s_{pqr}^{D_{pqr}} + x'_{pqr} + 1)}{(n+3)(s_{pqr}^{D_{pqr}} + 1)}, \quad we \ \ obtain,$$

$$C_{pqr}^{D'_{pqr}} = C_{pqr}^{D_{pqr}} e^{\frac{\tau_{pqr} - P(x_{pqr}=1|D_{pqr})}{m_{pqr}}\left\{\frac{(n+2)(s_{pqr}^{D_{pqr}} + x'_{pqr}+1)}{(n+3)(s_{pqr}^{D_{pqr}}+1)} - 1\right\}}$$

$$= C_{pqr}^{D_{pqr}} e^{\frac{\tau_{pqr} - P(x_{pqr}=1|D_{pqr})}{m_{pqr}}\left\{\frac{(n+2)x'_{pqr}}{s_{pqr}^{D_{pqr}}+1} - 1\right\}}$$

$$= C_{pqr}^{D_{pqr}} e^{\frac{\tau_{pqr} - P(x_{pqr}=1|D_{pqr})}{m_{pqr}}\left\{\frac{x'_{pqr}}{P(x_{pqr}=1|D_{pqr})} - 1\right\}} \quad \left[\ since \ \ P(x_{pqr} = 1|D_{pqr}) = \frac{s_{pqr}^{D_{pqr}} + 1}{n+2} \right]$$

Case 2: $\tau_{pqr} < P(x_{pqr} = 1|D_{pqr})$ and $\tau_{pqr} > P(x_{pqr} = 1|D'_{pqr})$
It is straightforward to see,

$$C_{pqr}^{D'_{pqr}} = w_{pqr} e^{\frac{\tau_{pqr} - P(x_{pqr}=1|D_{pqr})}{m_{pqr}}\left\{\frac{x'_{pqr}}{P(x_{pqr}=1|D_{pqr})}\right\}}$$

These two cases can be expressed together as,

$$C_{pqr}^{D'_{pqr}} = \left(I_{\tau_{pqr}}^{D_{pqr}} C_{pqr}^{D_{pqr}} + (1 - I_{\tau_{pqr}}^{D_{pqr}}) w_{pqr}\right) e^{\frac{\tau_{pqr} - P(x_{pqr}=1|D_{pqr})}{m_{pqr}}\left\{\frac{x'_{pqr}}{P(x_{pqr}=1|D_{pqr})} - I_{\tau_{pqr}}^{D_{pqr}}\right\}}$$

where $I_{\tau_{pqr}}^{D_{pqr}}$ is an indicator variable which takes value 1 if $\tau_{pqr} > P(x_{pqr} = 1|D_{pqr})$, else 0. Then, the overall cost considering all the factors, in accordance with (5), is given by (6).

We note that C_{pq}^{init} takes into consideration the influence of factors prevalent at the different alternatives on the effective costs. Similarly, costs involved among alternatives at successive stages (for instance, due to transport, tax, and handling of inventory in transaction etc.) can also be incorporated, taking into account uncertainty as gathered from historical data. The value of weights assigned to the different categories may be suitably adjusted for analyzing the overall cost across disparate supply chain application domains.

3.2 Modeling the Impact of Acquaintances and Distances

The impact of previous experience as a result of relationships among the different entities (e.g. suppliers/consumers at successive stages) is another important factor that has been overlooked thus far in the literature: if the experience is fruitful, the entities are likely to transact together as a part of supply chain again. In fact, this behavior is even more pronounced in case of global supply chains as the experience, between entities at successive alternatives, percolates down the supply chain. Furthermore, the distance dimensions also play a crucial role in the formation of global supply chains. In [8], these distances have been characterized into the following categories: cultural (e.g. religion, race, social norms, language), administrative and political (e.g. colony-colonizer links, currencies, trading arrangements), geographic (e.g. climate, waterway access, transportation and communication links, physical remoteness), and economic (e.g. information/knowledge, costs and quality of natural, financial and human resources, different consumer incomes)

An important observation is in order. These distances are a function of a pair of disparate alternatives at successive stages rather than being dependent on a single alternative. Thus, the whole process of the supply chain formation can be analyzed by using the following model:

1. Each of the alternatives is represented by a node.
2. For each node S_{pq}, a probability value $P_{pq} = \mathrm{P}(S_{pq})$ is calculated using (6) from Theorem 1 as
$$P_{pq} = \frac{\sum_{j=1}^{|S_p|} C_{pj}^{D''_{pj}} - C_{pq}^{D''_{pq}}}{(|S_p| - 1)\sum_{j=1}^{|S_p|} C_{pj}^{D''_{pj}}} \tag{7}$$
3. The impact of acquaintance between alternatives, S_{ij} and $S_{l,k}$, $i \in \{1, 2, \ldots, N-1\}$, $l = i+1$, $j \in \{1, 2, \ldots, |S_i|\}$, $k \in \{1, 2, \ldots, |S_{i+1}|\}$, on SCF is reflected by the corresponding acquaintance edge having probability,
$$P_{ijlk}^{AB} = \frac{AB_{ijlk}}{\sum_{j=1}^{|E_{ij}|} AB_{ijlk}}$$
where, AB_{ijlk} denotes acquaintance benefit of alternative k at stage l; and E_{ij}^{AB} is the set of acquaintance edges that are outbound from alternative j at stage i.
4. The impact of distance between alternatives, S_{ij} and S_{lk}, $i \in \{1, 2, \ldots, N-1\}$, $l = i+1$, $j \in \{1, 2, \ldots, |S_i|\}$, $k \in \{1, 2, \ldots, |S_l|\}$, on SCF is reflected by a corresponding distance edge having probability,
$$P_{ijlk}^{DC} = \frac{\sum_{j=1}^{|E_{ij}^{DC}|} DC_{ijlk} - DC_{ijlk}}{(|E_{ij}^{DC}| - 1)\sum_{j=1}^{|E_{ij}^{DC}|} DC_{ijlk}}$$
with $DC_{ijlk} = W_C f_C(DC_{ijlk}^C) + W_A f_A(DC_{ijlk}^A) + W_G f_G(DC_{ijlk}^G) + W_E f_E(DC_{ijlk}^E)$; where W_C, W_A, W_G, W_E are non-negative weights indicating the importance of the different dimensions corresponding to cultural, administrative, geographic, and economic distance respectively; $f_C, f_A, f_G, f_E : \Re^+ \rightarrow \Re^+$ are monotonically increasing functions that map the respective distance values

to their equivalent perceived costs; and E_{ij}^{DC} is the set of distance edges that are outbound from alternative j at stage i.

5. The *acquaintance edge* and the *distance edge* between every pair of nodes in the underlying model are replaced by a single edge called the *influence edge* (with same orientation as the acquaintance edge) between the same nodes. The probability on this edge is given by,

$$P_{ijlk} = W * P_{ijlk}^{AB} + (1 - W) * P_{ijlk}^{DC},\ 0 \leq W \leq 1$$

(where W indicates a relative preference for acquaintance over distance.)

$$= W \frac{AB_{ijlk}}{\sum_{j=1}^{|E_{ij}|} AB_{ijlk}} + (1 - W) \frac{\sum_{j=1}^{|E_{ij}^{DC}|} DC_{ijlk} - DC_{ijlk}}{(|E_{ij}^{DC}| - 1) \sum_{j=1}^{|E_{ij}^{DC}|} DC_{ijlk}} \tag{8}$$

Note that (8) is a valid probability measure since the sum of probabilities on all influence edges equals 1. Additionally, defining the probabilities this way is intuitive since the greater the acquaintance and the lesser the distance between two particular alternatives at successive stages is, the more likely the possibility of these alternatives being aligned again in a supply chain is.

4 Explaining the Dynamics of Supply Chain Formation

The dynamics of supply chain formation can be elegantly enunciated by using the following algorithm, based on the EcoSupply Model:

1. For each alternative q in stage p, draw a node with a probability value P_{pq} computed using (7).
2. Define the edge probabilities, for every pair of nodes representing alternatives at successive stages, using (8).
3. Add a dummy node, *Start*, which represents the stage S_0, and outbound edges to every node in S_1, with probability on each edge set to $\frac{1}{|S_1|}$. Further, set the probability value at *Start* to 1. (Note that this node serves the purpose of modeling multiple sources in the supply chain, which is another issue that has not been addressed in the literature thus far.)
4. Add a dummy node, *End*, which represents the stage S_{N+1}, and inbound edges from every node in S_N, with probability on each edge set to 1. Further, set the probability value at *End* to 1.
5. The probability of formation of a particular supply chain, $SC = A_1 A_2 \ldots A_N$, with $A_i\ i \in \{1, 2, \ldots, N\}$ denoting the alternative chosen at the stage i, is given by,

$$P_{SC} = P_{StartA_1} \prod_{i=1}^{N} P_{A_i} P_{A_i A_{i+1}} \tag{9}$$

The algorithm considers all the factors, depending on a particular alternative or a pair of alternatives at adjacent stages, that define the ecosystem: the probability on the nodes indicates the influence of factors restricted to a location whereas the probability on the edges indicates the influence of factors governing more than a single location. A change in any of these factors results in change in the probability values, given by (7) and (8), and a corresponding change in the probability of formation of an end to end supply chain, as indicated by (9). A relatively favorable ecosystem at an alternative, with respect to other alternatives, results in an increase in the corresponding probability of that alternative being a preferred choice for its stage, in the end-to-end supply chain.

5 Summary and Future Work

Modeling the impact of ecosystem on the supply chain formation is a topic of immense significance and has wide practical implications. Factors such as tax constraints and inventory handling costs have been well studied in the literature; however, several other important considerations such as the economic and cultural heterogeneity that constitute the entire ecosystem have been conspicuously ignored. In this paper, we proposed a generic Bayesian framework, EcoSupply, to model the dynamics of the supply chain formation. Specifically, we have illustrated how a change in the ecosystem accompanies a change in the local business alignments, and thereby the global supply chain dynamics. We have also showed how the acquaintances among the stakeholders greatly influence the future decisions regarding their collaboration. An important future direction would be to apply the EcoSupply model in different domains, for instance, food industry, automobile industry, financial sector, etc.

References

1. Walsh, W.E., Wellman, M.P., Ygge, F.: Combinatorial auctions for supply chain formation. In: Proceedings of the 2nd ACM Conference on Electronic Commerce (EC), pp. 260–269 (2000)
2. Parkes, D.C.: iBundle: An efficient ascending price bundle auction. In: ACM Conference on Electronic Commerce, pp. 148–157 (1999)
3. Antras, P., Helpman, E.: Global sourcing. J. Polit. Econ. 112, 552–580 (2004)
4. Grossman, G.M., Helpman, E.: Outsourcing in a global economy. Rev. Econ. Stud. 72(1), 135–159 (2005)
5. Goetschalckx, M., Vidal, C.J., Dogan, K.: Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. Eur. J. Oper. Res. 143(1), 1–18 (2002)
6. Oh, H.C., Karimi, I.A.: Regulatory factors and capacity-expansion planning in global chemical supply chains. Ind. Eng. Chem. Res. 43, 3364–3380 (2004)
7. Balaji, K., Viswanadham, N.: A tax integrated approach for global supply chain network planning. IEEE Trans. Auto. Sc. and Engg. (T-ASE) 5(4), 587–596 (2008)
8. Ghemawat, P.: Distance Still Matters. Harvard Business Review (2001)
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, Hoboken (2007)