A MODEL OF COMMUNICATION AS A RANDOM WALK ON THE SEMANTIC TREE

Vladislav B. Kovchegov
vlad_kovchegov@yahoo.com

Abstract

This paper contains a few models on free communications (communications without agendas). We will describe two types of free communication models. The first model is a model about communications with a given theme and it was implemented as a Markov stochastic process defined on the set of partitions. The second type of model involves not only partitions, but partitions and themes as well. This means that the conversation process is defined on the marking partitions and this looks like a random walk on the semantic trees.

Key words: *social network, communication, random walk, semantic tree.*

The common interests and three levels of communications. We first have to define what common interests mean in a given model. If we have two words a and b and there exists a common prefix c, then a=c* and b=c*, where symbol * means the same set of letters (for an arbitrary alphabet). If person one's favorite set contains a and does not contain b and the second person's favorite set b and does not contain a, we can say that person a has common interests on level c or c-interest. We will differentiate three levels of communications: level one or the general level, level two or real common interests, when people are emotionally involved in the same business, and level three or private, heart-to-heart communications. Second level communication generates (sometimes) deep positive or negative relations. These types of communications are the basis for the emergence of new groups and the crash of old ones. The third level of communication leads to love affairs and/or the emergence or destruction of families. For this type of communication partners first use body language, the language of glances and touch.

The free verbal and nonverbal communication model. The free communication model is the model of a situation when participants do not have an agenda. We describe two types of free communication models. The first type is the situation where participants sit in a given place. The second situation is the model of completely free communication, where participants do not have an agenda and move freely. The full description of the free conversation model for the case where people stay at a given location involves a more sophisticate mathematical language.

We start our description from the **first type** of communication model. Suppose we have a set of people I and a neighbor function N. The set of subsets of I is called a partition of I if the union of both sets is the set I and the intersection of the two is the empty set. We call a partition of I a partition agreed with a neighbor's function N if every element of the partition is a proper subset of at least one N(j) for the same j. Let us denote the set partition of I that agrees with a neighbor's function N by symbol H(I).

We can similar find the conditional probability for theme t(j,k) to be accepted as the suitable theme for conversation given as: P{t(j,k)| j } =v (|sub(t(j,k),N(j))| - 1)w(j,k)/|F(j)|, where theme t(j,k) belongs to F(j), has number k, and | F(j)| is the number of elements in set F(j). Symbol sub(t(j,k), N(j)) is the set of people from set N(J) that support theme t(j,k). We will later explain how to find the coefficient of v.

We need to, however, define the random process for the set of marking partitions agreed with neighbor function N. Any element of the partition will be marked by theme type and time. We use symbol Null to represent silence. Like the theme of "Weather", Null (or Silence) is a universal neutral theme. When a person stays silent it means that they are doing something: breathing, listening, eating, chewing, swallowing, watching, looking, sniffing, moving something, writing, reading, sleeping, walking, running, thinking, doing a job, and so on. Coefficient v must be found from the condition: sum of all P{t | j} for all themes t from F(j) and Null equals one. So if we want additional details we have to substitute actions instead of Null. When a local conversation for the same subgroup (element of partition) is over, all members get the same theme of Null and start a new time. Everyone in this group became a subset and new element of the partition. For instance, we have a conversation subgroup B={1, 5, 9} marked by theme b and time z. Suppose, that when the time of conversation equals 23 minutes the conversation ends. We need to transform subgroup B into three subgroups {1}, {2}, {9} and mark all three by theme Null. The new local time must be started, too.

The partitions, the marked partitions and transitions . We will start this item from example.

Example. Suppose we have I={1, 2, 3, 4} and N(1)={1, 2, 4}, N(2)={2, 1, 3 }, N(3)={3, 2, 4}, N(4)={4, 1, 3 }. Let us assume we have the second level communication and suppose that the favorite set of themes F(j) has an empty intersection for all sets and a nonempty intersection only for every distinct two. In this case partition D = ({1, 2}, {4, 3}) is the partition agreed with a neighbor's function N. The marked partition is partition D_m = ({1, 2}, t(1,k); {4, 3}, t(3,s)), where t(1,k) belongs to F(1) and t(3,s) belongs to F(3). This means that the first conversation subgroup {1, 2} discusses theme t(1,k) and the second subgroup {4, 3} discusses theme t(3,s). Let us infer that in random time z the second team stops talking, but the first team stays talking. This signifies that the initial marked partition will be transformed into the next one: ({1, 2}, t(1,k); {4, 3}, t(3,s)) → ({1, 2}, t(1,k); {4}, Null; { 3},Null). Then there are few variants: ({1, 2}, t(1,k); {4, 3}, t(3,s)) → ({1, 2}, t(1,k); {4}, Null; {3},Null) → ({1}, Null; {2}, Null; {4}, Null; {3}, Null) → ({1, 4}, t(4,k); {2}, Null; { 3},Null) and so on. We can not use, for instance, partition ({1, 2, 3}, {4}) because this partition does not agree with the neighbor's function N (element { 1, 2, 3 } is not a part of N(j), for all j). Nevertheless, in this case it is easy to describe all set H(I) (the set partition of I that agrees with neighbor function N): H(I) = { ({1}, {2}, {3}, {4}), ({1, 2}, {3, 4}), ({1, 2}, {3}, {4}), ({1, 4}, {3, 2}), ({1, 4}, {3}, { 2}), ({1, 2, 4}, {3}), ({2, 3, 4}, {1}), ({1, 3, 4}, {2}), ({2, 3, 1}, {4}) }.

We thus have only 11 states and all of them are not fit for modelling the second level of conversation. Instance states ({1, 2, 4}, {3}), ({2, 3, 4}, {1}), ({1, 3, 4}, {2}), ({2, 3, 1}, {4}) can be fit for level one conversation. We deduce from our assumption: any three people do not have common interests. So for the modelling of solely the second level of communication we have to reject these partitions.

However, there does not exist pure communications in real life. More often than not real communication is a mix of all three levels. This means that free communication generates a process that emerges (clashes) groups, families, and so on. What does this transition mean? The transition ({1, 2}, t(1,k); {3}, Null; {4}, Null) → ({1, 2, 3}, t(1,s); {4}, Null) means that person 3 joins conversation team {1, 2}. Why is this possible? This occurs because team {1, 2} changes its type of theme: they skip theme t(1,k)=crd and discuss theme t(1,s)=c, where c is level one's theme and c, r, and d are words in a theme's alphabet. In this model the semantic field/scale contains alphabets (general letters, special letters) and sets of words. General letters are letters that code the general themes (the first level of communication). A set of words has the structure (general words, special words) and grammar. For example, "letter" Rm would code the general theme "Rumor" and "letter" Sp represents the theme "Sport." It then follows that word RmSp represents the theme: a rumor in sport. If we use the theme "Football," "letter" Fo codes this and theme "Basketball" is coded by "letter" Ba and words SpFo represent football. Suppose that "letter" Gm represents theme "Game." We can find at least two people who like to talk about sport, but differentiate on the type of sport. So, if t(1,k)=SpFoGm and t({1, 2}, s)=RmSpFo, we can interpret this transition as the transition for the theme of talk ranging from questions about football to rumors about players, coaches, and so on. In this case the special word is transformed into the general word by adding one "letter" Rm. But rumors aside, only one letter transfers the special theme (special word) into the general theme (general word). We define the special word as a word in a special alphabet. The general word is a word with at least one letter from a general alphabet. This kind of explanation suggests to us that transition ({ 1, 2}, t(1,k); { 3}, Null; { 4} , Null) → ({ 1, 2 }, t(1,s); {3} , Null; { 4 }, Null) occurs before transition ({ 1, 2}, t(1,k); { 3}, Null; { 4}, Null) → ({ 1, 2, 3 }, t(1,s); { 4 }, Null). Finally, we have the chain of transitions ({1, 2 }, t(1,k); { 3 }, Null; {4 }, Null) → ({1, 2 }, t(1,s); { 3 }, Null; {4 }, Null)) → ({1, 2, 3} , t(1,s); {4} , Null). We must only be concerned about the moment in time when there emerges a new theme that changes the structure of interests or sours future alterations. If we want to create a completely free communication model where participants do not have an agenda and are able to move about, we have to use as a state the arbitrary partition of I. The system then stays in some state (partition agreed with a neighbor's function N) for a random period of time and can randomly change its state. If the time between transitions has an exponential distribution we have a Poisson process. Let us define the probability of emergence of conversation subgroup A for a small period of time T by formula Pr{emergence of conversation subgroup A} = m(A)T + o(T) where A is a subset of N(j), for the same j, $|A| \geq 2$ and the probability equals zero otherwise. Then if the intersection of all favorite subsets F for people from A is empty the theme for conversation is gotten from the general set G (level one conversation). Otherwise a person of subset A chooses the conversation theme from the intersection. The probability of collapse for a small period of time T equals M(A)T + o(T). So the probability for conversation group A to emerge with theme t equals (m(A)T + o(T))(q(j)p(t|j) + ... + q(k)p(t|k)), where q(j) is the probability that person j initiates the conversation theme, p(t|j) is the conditional probability that theme t is chosen by person j and A={j, ... ,k }. For the case of four people we do not have a large number of states or partitions. The first partition agrees with a neighbor's function N (and N(1)={ 1, 2, 4}, N(2)={ 2, 1, 4}, N(3)={ 3, 2, 4}, N(4)={ 4, 1, 3}) is D=({ 1},Null; { 2}, Null; { 3}, Null; { 4}, Null). The second group of marked partitions are partitions D(1,2;t)= ({ 1, 2}, t; { 3}, Null; { 4}, Null), D(1,3;t)= ({ 1, 3}, t; { 1}, Null; { 3}, Null), D(2,3;t)= ({ 3, 2}, t; { 1}, Null; { 4}, Null), D(1,4;t)= ({ 1, 4}, t; { 3}, Null; { 2}, Null), D(2,4;t)= ({ 4, 2}, t; { 3}, Null; { 1}, Null), D(3,4;t)= ({ 3, 4}, t; { 1}, Null; { 2}, Null), where conversation theme t belongs to the participants intersection of their favorite set or to the set of general conversation themes. The third group of state is the group of partitions D(k,j;t1,t2) = ({ k, j}, t1; { s, r}, t2), for all different combinations of { k, j, s, r}, where k, j, s, r belongs to the set I={ 1, 2, 3, 4} and { k, j }={ j, k, }, { r, s }={ s, r}, and t1 and t2 are the conversation themes. The last group of marked partitions is the group of partitions D(j;t) = ({ k, s, r}, t; { j}, Null), for different j, where { k, s, r} = I \ { j}, and t is the theme. Symbol Null does not represents spiking actions (for instance, not verbal communication actions).

We are now ready to write the equation for the first type of free communication model under one assumption: **the time between transition has exponential distribution.** It may not be a realistic assumption (we achieve a pure Markov process), but we can at least finish the solution successfully. In the case of a semi-Markov process we get a lot of problems with analysis. If we do not care about themes we get the system of thirteen differential equations. We have a set of favorite themes F(j), where j belongs to I, and a set of general themes G (every person j may have the favorite set of general themes G(j)). In this case we get 1 + 3(|F(1) F(2)| + |F(1) F(3)| + |F(1) F(4)| + |F(2) F(3)| + |F(2) F(4)| + |F(3) F(4)|) + (| F(1) F(2) F(3)| + (| F(1) F(2) F(3)| + (| F(1) F(2) F(3)| + (| F(1) F(2) F(3)|) + 16 | G | differential equations.

We have divided the communication process on the two independent parts. The first part (model 1) is dedicated to process that emerges the communication groups (having verbal conversation groups first). The second part (model 2) of the communication problem is the stochastic model of theme flow in every subgroup, when the communication groups are formed.

Model 1. Emerge the communications subgroups that discusses one theme. We ignore the type of theme discussed. Let D_1 = D, D_2 =

D(1;T) = ({ 1}, Null; {2, 3, 4}), D_3 = D(2;t), D_4 = D(3,t), D_5 = D(4,t), D_6 = D(1,2;t), D_7 = D(3,4; t), D_8 = D(1, 4; t), D_9 = D(2,3;t), D_{10} = D(2,4; t), D_{11} = D(1, 3; t), D_{12} = D(1, 2;t1,t2), D_{13} = D(1, 3; t1, t2), D_{14} = D(1,4; t1, t2), for all themes t, t_1, t_2.

Right now we will describe transition graph for 14 states. The transition graph can be define by set { 1 $\leftrightarrow$ 2, 1 $\leftrightarrow$ 3, 1 $\leftrightarrow$ 4, 1 $\leftrightarrow$ 5, 6 $\leftrightarrow$ 12, 12 $\leftrightarrow$ 7, 7 $\leftrightarrow$ 1, 1 $\leftrightarrow$ 8, 8 $\leftrightarrow$ 13, 13 $\leftrightarrow$ 9, 9 $\leftrightarrow$ 1, 1 $\leftrightarrow$ 11, 11 $\leftrightarrow$ 14, 14 $\leftrightarrow$ 10, 10 $\leftrightarrow$ 1 }, where k $\leftrightarrow$ j means that exist transition k $\rightarrow$ j (with probability Pr(X(T+s)=k/X(T)=j) = m(k,j)s + o(s)) and transition j $\rightarrow$ k (with probability Pr(X(T+s)=j/X(T)=k) = m(j,k)s + o(s)) . Let X(T) be our stochastic process and p_j(T) = Pr{ X(T) = D_j } , where j= 1, . . . ,14.

The system's differential equation that describes the evolution of initial distribution is:

$$dp_1(T)/dT = (-m(1,2) - \ldots - m(1,11))p_1 + m(2,1)p_2 + \ldots + m(11,1)p_{11}$$

$$dp_2(T)/dT = -m(2,1)p_2 + m(1,2)p_1$$

$$dp_3(T)/dT = -m(3,1)p_2 + m(1,3)p_1$$

$$dp_4(T)/dT = -m(4,1)p_2 + m(1,4)p_1$$

$$dp_5(T)/dT = -m(5,1)p_2 + m(1,5)p_1$$

$$dp_6(T)/dT = -(m(6,1) + m(6,12))p_6 + m(12,6)p_{12} + m(1,6)p_1$$

$$dp_7(T)/dT = -(m(7,1) + m(7,12))p_7 + m(12,7)p_{12} + m(1,7)p_1$$

$$dp_8(T)/dT = -(m(8,1) + m(8,13))p_8 + m(13,8)p_{13} + m(1,8)p_1$$

$$dp_9(T)/dT = -(m(9,1) + m(9,13))p_9 + m(13,9)p_{13} + m(1,9)p_1$$

$$dp_{10}(T)/dT = -(m(10,1) + m(10,14))p_{10} + m(14,10)p_{14} + m(1,10)p_1$$

$$dp_{11}(T)/dT = -(m(11,1) + m(11,14))p_{11} + m(14,11)p_{14} + m(1,11)p_1$$

$$dp_{12}(T)/dT = -(m(12,6) + m(12,7))p_{12} + m(6,12)p_6 + m(7,12)p_7$$

$$dp_{13}(T)/dT = -(m(13,8) + m(13,9))p_{13} + m(8,13)p_8 + m(9,13)p_9$$

$$dp_{14}(T)/dT = -(m(14,10) + m(14,11))p_{14} + m(10,14)p_{10} + m(11,14)p_{11}$$

where Pr{X(T+s) = D_k /X(T)=D_j }=m(j,k)s + o(s), for all j and k . It is easy to find the stationary points (measure) of the system. We can represent the system of equation in the matrix form as: dp(T)/dT = Ap(T) where p(T) is vector of probabilities and A is a matrix of coefficients. In the case where all coefficients of intensity m(k,j) =a, for all k and j, where a is a positive constant our system of equation looks like: dp(T)/dT = aAp(T) We can now find the exact solution of the system. For this purpose we will find the eigen values for matrix A. Matrix A has a negative vector of the eigen values

m = (e(1), e(2), e(3), e(4), e(5), e(6), e(7), e(8), e(9), e(10), e(11), e(12), e(13), e(14)) = (0, A/3 + 29/A -5, -A/6 -29/2A -5 + 0.5/($\sqrt{3}$(A/3 -29/A)),

-A/6 29/2A -5 -0.5/($\sqrt{3}$(A/3 -29/A)), -2+ $\sqrt{2}$, -2- $\sqrt{2}$, -2+ $\sqrt{2}$, -2 - $\sqrt{2}$, -1, -2,-1, -2, -1, -2) , where A=$(-648 + 3/\sqrt{26511})^{\frac{1}{3}}$.

So, our system of linear equations has the solution (see, for instance, [1])

p(T) = $\exp(e(1)T)Z_1$ + $\exp(e(2)T)Z_2$ + $\exp(e(3)T)Z_3$ + $\exp(e(4)T)Z_4$ + $\exp(e(5)T)Z_5$ + $\exp(e(6))Z_6$ + $\exp(e(7)T)Z_7$ + $\exp(e(8)T)Z_8$ + $\exp(e(9)T)Z_9$ + $\exp(e(10)T)Z_10$ + $\exp(e(11)T)Z_{11}$ + $\exp(e(12)T)Z_{12}$ + $\exp(e(13)T)Z_{13}$ + $\exp(e(14)Z_{14}$,

where matrices $Z_1, \dots, Z_{14}$ satisfy conditions Z_iZ_j =0,I = Z_1+ ... + Z_{14}, $Z_iZ_i = Z_i$, for all j and i from the set { 1, 2, ... ,14 }, I is identical matrix, and T is time.

Model 2. Communication as random walk on the semantic tree. We can now describe the second part of the communication process as stochastic process on the set marked partitions. For this, we need additional information about the person. We will then use the semantic field and body language. Suppose that A is the alphabet for the semantic field (the theme language). The arbitrary word (every word represents a theme or a set of themes) is the sequence of letters. For instance, a=S1H2K5 is word for the alphabet A= { S1, H2, K5 }. The conversation process builds or destroys random processes. The typical trajectory of the random process looks like the sequence X(T0) =G2, X(T1) =AB, X(T2) = ABC, X(T3) = ABCD, X(T4)= ABCDF, X(T5) = G1, X(T6) =ABC, X(T7) = ABCKL, X(T8) =G3, etc. where T0, T1, T2, are random moments in time. This means that the conversation process starts from the general theme (G2) and proceeds onto special questions. We can see a growing deepness in conversation (T1-T5) and at moment T5 the conversation process goes back to the general area again, and so on. How can we formalize this type of process? We use the randomized formal grammar method. What do mean by this? Let us propose that the participants have a non-empty set of common favorite themes. For instance, suppose that theme b= ABCDF is a common favorite theme for all members of the conversation team. Someone initiates the following approach: she or he offers theme AB. Someone then makes the next approach (theme ABC) and so on. In a few steps the conversation subgroup reaches the desirable theme ABCDF. When participants tire or the theme is exhausted the process moves to the opposite direction. Our next question is how do different evaluations and/or opinions about events (inside theme ABCDF) transform relations (see 3rd step). We represent the semantic scale as trees: S-trees for special topics, G-trees for general topics (funny stories, jokes, anecdotes, rumors, and C-trees for current events. Suppose we the alphabet of themes A and suppose that any nodes on the tree are marked by letters from the alphabet and by a set of events. The roots are marked by symbol S for S-tree, symbol G for G-tree, and symbol C for C-tree. Now let us assume that all three symbols do not belong to the alphabet. The nodes on the tree represent words. How do we find these words? We take the shortest path from the root to a given node and write down (from left to right) all letters that lie on the path from the root to given node. The favorite or sick sets of themes are the set of nodes on the G - or S - trees. In this case the conversation process is a random walk on the trees.

We can combine all trees into one by adding one additional extra root with three edges to the roots of S-, G-, and C- trees. How do we calculate the probability of coming from a given node to the next neighbor or to jump to another place? Suppose we have node j on the semantic tree. For every participant we calculate the set of shortest paths from a given node to all elements of the favorite set. For a given person we must calculate the number of paths that go from a given node j to a member of the favorite set with the nearest edges of node j (this means that the edge is a part of the trajectory). The probability of going onto the next node must be proportional to its number. For instance, if node j has the set of neighbors { k, l, r } on the semantic tree with a number of paths that start at j, that go throughout k (or l or r) equaling to M(j,k) (or M(j,l) or M(j,r)), this probability walk to k must be proportional M(j,k) and so on.

We can now demonstrate an example of a conversation model (random walk on the tree model). Suppose we have alphabet A and a semantic tree for two people. For our example we will use very simple alphabet A = { Cu, D1, D2, Sp, Fu, Ba, Ar, Mu, Pa}, where "letter" Cu (number 2) present Cooking, D1 (number 5) and D2 (number 6) are some dishes. The letter Sp (number 3) is abbreviation for Sport, Fu (number 7) present Football and Ba (number 8) is Basketball. The "letter" Ar (number 4) present Art and Mu (number 9) present Music, Pa (number 10) present Performance. Let us use symbol S (number 1) as initial symbol. The semantic tree we be defined by set { 1 $\rightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4, 2 $\rightarrow$ 5, 2 $\rightarrow$ 6, 3 $\rightarrow$ 7, 3 $\rightarrow$ 8, 4 $\rightarrow$ 9, 4 $\rightarrow$ 10 }. Having case one of six partitions D(1,2) = ({1, 2}, {3}, {4}), D(1,3), D(1,4), D(2,3), D(2,4), D(3,4)) and case D(2). The favorite sets are F_1={Cu, D1, Sp, Ba} or {2, 5, 3, 8} and F_2={Sp, D2, Fu, Ra } or { 3, 6, 7, 8 } .

This signifies that we will use a number theme instead of a word theme (i.e. number 2 instead of Cu (Cooking)). The arrow and number on the arrow denotes the probabilities of skipping onto the next position. We have the graph of transition and we can therefore write the system of differential equations. Let us denote the conditional probability Pr{X(T)=D(A,t) / X(T)=D(A)} by symbol p(t,A;T), where t represents the theme for conversation group A (suppose A= {1, 2 } or {1, 3} or {1, 4} or {2, 4}).

The system's differential equation that describes the evolution of the initial distribution on the semantic tree is: (Suppose D(A) is D(1,2) =({1, 2}, {3}, {4}, F_1 = {2, 5, 3, 8}, F_2= { 3, 6, 7, 3, 8}, Pr{X(T+s) = D(A,t) /X(T)=D(A)}=n(j,k)s + o(s), n(j,k) > 0 and let p(t,T) = p(t,{1,2};T))

$$dp(1,T)/dT = -(n(1,2) + n(1,3) + n(1,4))p(1,T) + n(2,1)p(2,T)+$$
$$n(3,1)p(3,T) + n(4,1)p(4,T)$$
$$dp(2,T)/dT = -(n(2,1) + n(2,6))p(2,T) + n(1,2)p(1,T)+$$
$$n(5,2)p(5,T) + n(6,2)p(6,T)$$
$$dp(3,T)/dT = -(n(3,1) + n(3,8))p(3,T) + n(1,3)p(1,T)+$$
$$n(7,3)p(7,T) + n(8,3)p(8,T)$$

$$dp(4,T)/dT = -(n(4,1) + n(4,9))p(4,T) + n(1,4)p(1,T)+$$
$$n(9,4)p(9,T) + n(10,4)p(10,T)$$
$$dp(6,T)/dT = -n(6,2)p(6,T) + n(2,6)p(2,T)$$
$$dp(8,T)/dT = -n(8,3)p(8,T) + n(3,8)p(3,T)$$
$$dp(9,T)/dT = -n(9,4)p(9,T) + n(4,9)p(4,T)$$
$$dp(5,T)/dT = -n(5,2)p(5,T)$$
$$dp(7,T)/dT = -n(7,3)p(7,T)$$
$$dp(10,T)/dT = -n(10,4)p(10,T)$$

, where p(1,T) + p(2,T) + p(3,T) + p(4,T) + p(5,T) + p(6,T) + p(7,T) + p(8,T) + p(9,T) + p(10,T) =1

It is easy to find the stationary points (measure p(k,T)=p(k), for all T) of the system and we can additionally find the system's solution. We need to write similar equations for all partitions and then plan (at a later point) to also use the "natural attractions." This attractiveness is a very important property and it known as the relational property.

Conclusion. We can similarly make the model of conflict. For the modelling of a conflict situation we have to just add the set of "sick" theme for any person (where a "sick" theme is a theme that invokes unusual reactions). Conflicts occur when the marking partitions are reached that are marked by "sick" theme(s). The model of conflict pays more attention to details and therefore essentially generates more states of the system and large numbers of equations. If we want additional details we can use body language as well. For instance, in the third level of communication (greetings, heart-to-heart) body language is more important than oral. But for all of these models we need to write a large system of equations.

REFERENCES

1. Gantmacher F.R.1974, **The Theory of Matrices** Chelsea, New York, 1974