A Novel Exploration of Diffusion Process based on Multi-types Galton-Watson Forests

Yanjiao Zhu\textsuperscript{1}, Qilin Li\textsuperscript{2}, Wanquan Liu\textsuperscript{3*}, Chuancun Yin\textsuperscript{1}, Zhenlong Gao\textsuperscript{1}

\textsuperscript{1} School of Statistics, Qufu Normal University, Qufu, China
\textsuperscript{2} Department of Computing, Curtin University, Perth, Australia
\textsuperscript{3} School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China

Email: liuwq63@mail.sysu.edu.cn

Abstract

Diffusion is a commonly used technique for spreading information from point to point on a graph. The rationale behind diffusion is not clear. And the multi-types Galton-Watson forest is a random model of population growth without space or any other resource constraints. In this paper, we use the degenerated multi-types Galton-Watson forest (MGWF) to interpret the diffusion process and establish an equivalent relationship between them. With the two-phase setting of the MGWF, one can interpret the diffusion process and the Google PageRank system explicitly. It also improves the convergence behavior of the iterative diffusion process and Google PageRank system. We validate the proposal by experiment while providing new research directions.

Keywords: diffusion process, Galton-Watson forest, Google PageRank system

1. Introduction

The diffusion process is a phenomenon of spontaneous spreading of particles from a region of high concentration to one of the lower concentration regions. In other words, the diffusion process can be viewed as dynamic evolving when a system is not in equilibrium and random motion tends to bring its states towards uniformity. The diffusion processes have been used to model many physical, biological, engineering, economic, and social phenomena. Also, diffusion processes have been used in machine learning, because of its capacity in capturing the intrinsic manifold structure of data,
and they have been widely used in many fields such as information retrieval [2, 3, 4, 5, 6]; clustering [7, 8]; saliency detection [9, 10]; image segmentation [5, 34]; semi-supervised learning [6, 20].

Technically, diffusion processes generally start from a given $N \times N$ affinity matrix $A$, which relates $N$ different elements to each other. An undirected graph $G = (V, E)$ can be represented by an affinity matrix $A \in \mathbb{R}^{N \times N}$, where the graph $G = (V, E)$ consists of $N$ nodes $v_i \in V$, and edges $e_{ij} \in E$, that link nodes to each other. The edges between node $i$ and node $j$ can be weighted by their affinity value $A_{ij}$. The diffusion process can be interpreted as a Markov random walk on graph $G$, where the probability of walking from one node to another is governed by a transition matrix $P$, in which the relation between $A$ and $P$ is described as below. We first define a $N \times N$ diagonal matrix $D$ with $D_{ii} = \Sigma_{k=1}^{N} A_{ik}$, then the transition matrix $P$ can be written as follow:

$$P = D^{-1}A.$$ 

Note that $P$ is a row-stochastic matrix (rows sum up to 1). The diffusion process is to place a particle on the node of an undirected graph $G$, with the initial distribution $F_0$ which is an $N \times N$ matrix and contains the initial information for every node. In the existing literature on the diffusion process, $F_0$ was taken differently as $A$ (affinity matrix) [11], $I$ (identity matrix) [13], $P$ (transition matrix) [35]. In this paper, we let $F_0 = P$ unless otherwise specified. Each particle transfers the information according to the transition matrix $P$, then the distribution of the particles that we get from one iteration is $F_1 = F_0P$. Going through this process iteratively, the distribution of the particles will reach a stable state $\Pi$ which is the matrix of the left eigenvector of the transition matrix $P$ with corresponding eigenvalue 1. This stable state is also the intrinsic manifold structure description of data. The random walk on the graph which contains three nodes is shown in Figure 1.

![Figure 1: Random walk on the graph which contains three nodes.](image)

One important application for the diffusion process was the web PageRank [26], in which each website corresponds to one node in the graph; the
hyperlink between websites is represented by the edge in the graph. However, one typical problem is that the user sometimes browses the website by entering the URL rather than clicking on the hyperlink, which causes some influence on the PageRank system as prior information. The Google PageRank system [26] solved the problem by using the following algorithm:

\[ F_n = \alpha F_{n-1} P + (1 - \alpha)Y, \]

where \( Y \) represents the prior information of the web PageRank system. The Google PageRank system is an extension of the diffusion process.

In the literature, the diffusion process and its extensions were explained by the random walk model on the graph, but it has certain limitations. (1) We didn’t have a clear understanding of how each particle moves, and its engineering significance was not obvious; (2) The transition matrix \( P \) for the diffusion processes depends on the affinity matrix \( A \) of the graph, which limits the expansion of the diffusion process; (3) For the Google PageRank system, the matrix \( Y \) cannot be visually explained in the graph. To address these issues, this paper uses a special multi-type Galton-Watson forest (MGWF) model [23, 24] to explain the diffusion process from a different perspective.

The MGWF is a generalization of mono-type Galton-Watson forests that describe the genealogy of a population in which individuals are differentiated by types. The MGWF \( \{Z_n\} \) has a visual representation of a multi-type branching process with multiple ancestors, where \( Z_n = (Z_1, Z_2, \ldots, Z_N) \), \( Z_1 \) represents the number of the particle of 1-type, \( \ldots \), \( Z_N \) represents the number of the particle for the \( N \)-type, respectively. For a \( i \)-type parent, it lives for a unit of time until upon death, produces \( j_1 \) children of 1-type, \( \ldots \), \( j_N \) children of \( N \)-type according to the branching law \( P^i(j_1, j_2, \ldots, j_N) \). As a counting process, MGWF mainly considers the ultimate property of population size \( \{Z_n\} \). An example of 3-type GWF is shown in Figure 2(a).
(a) The 3-type GWF \( \{Z_n\} \). In the 1st generation the number of 1, 2, 3-type are 1,1,1, respectively.

(b) MGWF with degenerated branching law. A parent particle can only produce one offspring particle.

Figure 2: The 3-type GWF and 3-type GWF with degenerated branching law

An MGWF with a degenerated branching law is not the focus of the research in the field of branching processes. However, it is of our interest in this paper to use it to explain the diffusion process. The degenerated branching law is: for every \( i \in 1, 2, \cdots, N \).

\[
P^i(\{Z \in \mathbb{N}^N : \sum_{j=1}^{N} Z_j \neq 1\}) = 0
\]

That is, a parent can produce only one offspring. Essentially, the members of this forest are infinite direct trees as shown in Figure 2(b). The N points in the random walk correspond to the N types of particles in the MGWF. The random walk on the graph is pulled longitudinally into particles to produce offspring by this way, in which the concept of "level" comes up.

To use the MGWF to explain the diffusion process explicitly, we innovatively introduce the concept of particle mutation (in a unit of time, \( i \)-type particles produce \( j \)-type particles, and there is a chance that the \( j \)-type particle will mutate into \( k \)-type particles). Therefore, the random process can be divided into two phases - branching phase and mutation phase as shown in Figure 3. By dividing each step of the diffusion process (It corresponds to
the unit time in the MGWF) into two phases, one not only can understand
the diffusion process more clearly, but also it would inspire us to learn the
diffusion process in a general setting.

With this two-phase setting, we can further consider MGWF with immi-
gration [37]. The classical theory of immigration is to introduce immi-
grants to ensure that the entire forest would not extinct so that immigrant particles
enter the forest and produce offspring together with the original particles.
However, we only consider immigration particles as another option different
from the branching-mutation mechanism to produce offspring. The model
of MGWF with immigration and mutation can be used to explain the well-
known Google PageRank system, which is a successful extension of the dif-
fusion process. This will also inspire us to interpret the Google PageRank
system from a new perspective.

In summary, the main contributions of this article are as follows:

- By introducing the novel concept of mutation into the MGWF with the
degenerated branching law, each step of the diffusion process is divided
into two phases - the branching phase and the mutation phase. Such
explanation provides novel insight on the rationale of the diffusion pro-
cess from a longitudinal perspective, compared to the existing random
walk interpretation in the literature.

- In the existing diffusion process, the transfer matrix $P$ depends on the
affinity matrix $A$. But by looking at the diffusion process from the

Figure 3: MGWF with mutation. A unit of time is divided into two phases - branching
phase and mutation phases.
perspective of the MGWF, the $P$ and $A$ can be separated as unrelated parts, which is a benefit for the extension of a more general diffusion process.

- Within the framework of branching-and-mutation, we propose a variant of the diffusion process in which the mutation probability can change. It is a more general diffusion process that includes the original as a special case.

- We further extend the MGWF by introducing immigration, which can interpret the Google PageRank system well.

- Finally, two examples (Example 5.1, 5.2) of the variant of diffusion process in which mutation probability can be changed prove that the variant of diffusion process can effectively improve the convergence rate of iteration.

This paper is organized as follows. In the preliminary section, we introduce the MGWF model in detail; In Section 3 we introduce the concept of mutation into the MGWF with degenerated branching law and explain the diffusion process under our new model; also, We introduce the MGWF model with immigration and use it to explain the Google PageRank system in section 4; in section 5, we provide some examples of the diffusion process and the Google PageRank system respectively, and show that possible modifications in the setting can increase the rate of convergence in the extended diffusion process; We draw conclusions and point out potential future directions in section 6.

2. Preliminaries

2.1. Random walk and Branching process

First, let us look at the definition of the Markov chain.

**Definition 2.1. Markov chain** Let $\{X_n, n \geq 0\}$ be a sequence of discrete random variables. For states space $\mathcal{S}$, the $\{X_n, n \geq 0\}$ satisfying the Markov properties:

$$
P(X_{n+1} = j|X_n = i, \cdots, X_0 = i_0) = P(X_{n+1} = j|X_n = i) = P_{i,j}
$$

for all $n = 1, 2, 3 \cdots$ and states $i_0, \cdots, i, j \in \mathcal{S}$, is called a Markov chain with initial distribution $F_0$ and transition probability matrix $P$. A finite-state Markov chain is a Markov chain in which $\mathcal{S}$ is finite.
A very important special case of the Markov chain is the random walk on an undirected graph. The random walk picks a neighbor at each step randomly and moves to that neighbor. Hence, the transition matrix is $P = D^{-1}A$, where $A$ is a $N \times N$ affinity matrix of the graph, and $D$ is the diagonal matrix with $D_{ii} = \sum_{j=1}^{N}A_{ij}$.

The simplest type of a branching process is the Galton-Watson process (GW process) which is also a Markov chain. The branching process is one of the classical fields in applied probability and it can deal with a mathematical representation for the population changes in which their members can reproduce and die, subject to laws of chance. The particles may be of different types depending on their age, energy, position, or other factors. However, the type of all particles is the same in the GW process. Let us define a GW process as a Markov chain $\{Z_n\} = Z_0, Z_1, Z_2, \cdots$, where $Z_n$ is a random variable describing the population size at the $n$-th generation. The Markov property can be interpreted as: in the $n$-th generation, the $Z_n$ individual independently gives rise to some number of offsprings $\xi_{n+1,1}, \xi_{n+1,2}, \cdots, \xi_{n+1,Z_n}$ for the $(n+1)$-th generation. $\xi_{n+1,j}$ can be thought of as the number of members in the $(n+1)$-th generation which are offsprings of the $j$-th member of the $n$ generation. Note that $\{\xi_{n,j}, n \geq 1, j \geq 1\}$ are identically distributed (having a common distribution $\{p_k\}$) non-negative integer-valued random variables. Thus, the cumulative number produced for the $(n+1)$-th generation is

$$Z_{n+1} = \xi_{n+1,1} + \xi_{n+1,2} + \cdots + \xi_{n+1,Z_n}.$$

Thus the probability of any future behavior of the process, when its current state is known exactly, can not be altered by any additional knowledge concerning its past behavior.

Two examples of Markov chains are shown in Figure [4] one as random walks on an undirected graph and another as a GW process separately.
2.2. The multi-type Galton-Watson forest with degenerated branching law

In this section, we consider a multi-type GW process without the assumption that all particles are of the same type. Set a N-type GW process \( \{Z_n\}_{n \geq 0} \), with the initiated state as

\[ Z_0 = i = (i_1, \cdots, i_N) \]

where \( i_k \in \mathbb{Z}_+ = \{0, 1, 2, \cdots \} \), \( k = 1, \cdots, N \), \( i_k \) is the number of individuals of \( k \)-type in the first generation. For the process with \( Z_0 = i \), the number of \( j \)-type individuals in the \( n \)-th generation is

\[ Z^{(i)}_n = (Z^{(i)}_{n,1}, Z^{(i)}_{n,2}, \cdots, Z^{(i)}_{n,N}) \]

where \( Z^{(i)}_{n,j} \) represents the number of particles of \( j \)-type in the \( n \)-th generation. We assume that the parent \( i \), lives for a unit of time and, upon death, produces children of all types and according to the offspring distribution

\[ \{P^i(j)\} = \{P^{(i_1,j_2,\cdots,i_N)}(j_1, j_2, \cdots, j_N)\}, i_k, j_k \in \mathbb{N}, k = 1, 2, \cdots, N \]

and the independence of other individuals. In which \( \{P^{(i_1,j_2,\cdots,i_N)}(j_1, j_2, \cdots, j_N)\} \) is the probability that the parents produce \( j_1 \) children of 1-type, \( \cdots, j_N \) children of \( N \)-type, and the parents are \( i_1 \) particles of 1-type, \( \cdots, i_N \) particles of \( N \)-type. The case with the initial number of particles being multiple is vividly called a multi-type Galton-Watson forest (MGWF).

Let \( e_i, i = 1, 2, \cdots, N \) denote a N-dimensional row-vector with \( i \)th component as 1, and the remaining being zero.
Next, a special N-type Galton-Watson forest is considered. There are $N$ particles in the system at the time 0, and their types are $1, 2, \cdots, N$ respectively. Each particle is considered as an ancestor of a multi-type GW process, i.e.

$$Z_0 = e_i$$

We denote the initial state of the MGWF as an $i$-type particle. Each ancestor lives for a unit of time and may produce children of all types upon death according to the offspring distribution $\{P^{e_i}(e_j)\}$. For convenience, we abbreviate it as:

$$\{P^{e_i}(e_j)\} = \{P^i(j)\}, i, j = 1, 2, \cdots, N,$$

where $\Sigma_{j=1}^N P^i(j) = 1$, in other words, for every $i \in 1, 2, \cdots, N$

$$P^i(\{Z \in \mathbb{N}^m : \Sigma_{j=1}^N Z_j \neq 1\}) = 0$$

i.e., each parent produces only one offspring. In this way, the number of multi-type branching tree formed in the forest is $N$. From \cite{23, 24}, we know the system is the MGWF with degenerate branching law, i.e. every element in the forest is an infinite trees. However, here we still call this system a MGWF in this paper. The system with a $N \times N$ branching law matrix $P$

$$P = (P^i(j))_{i,j=1,2,\cdots,N}$$

where for any $i = 1, 2, \cdots, N$, we have $\Sigma_{j=1}^N P^i(j) = 1$. The difference between the MGWF and the MGWF with degenerated branching law is graphically illustrated in Figure \cite{22}.

In the existing literature \cite{23, 24, 25, 27, 28}, many MGWFs mainly discuss the convergence of their corresponding production process, which is the sequence of generations of the individuals of the MGWF visited in lexicographical order. It is easy to check that the height process fully characterizes the forest. In the literature, it is assumed that the branching law is immutable. In this paper, we mainly discuss the change of the branching law in some special cases.

**Example 2.1.** Let us first consider an example where there are three types of particles in the system, i.e. $N = 3$. The 3-type GW forest with the
degenerated branching law $P$ is shown in Figure 2(b). The branching law matrix is

\[
P_{3\times3} = \begin{pmatrix}
0 & 0.6250 & 0.3750 \\
0.3846 & 0 & 0.6154 \\
0.2727 & 0.7273 & 0
\end{pmatrix}.
\]

3. Mutations in the MGWF

3.1. The Mutations

Let us consider mutation in the framework of MGWF with degenerated branching law. “Mutation” is not a new concept in the literature of the MGWF [29, 30]. The original definition is that a mutation producing $i$-type is the birth event of an individual of $j$-type from an individual of any type $j \neq i$. In this paper, we propose a novel variant of mutation which can better conform to the actual situation, and more importantly, it provides us more insight into the rationale of the diffusion process by separating it into two phases: branching and mutation.

**Definition 3.1. (Mutation)** When mutation time is ignored, a unit of time is divided into two phases- the branching phase and the mutation phase.

(i) In the branching phase, an $i$-type particle produces a $j$-type particle according to the branching law $P^i(j)$.

(ii) In the mutation phase. The mutation occurs: the $j$-type particle may mutate into a $k$-type particle according to the probability of mutation $T^j(k)$, $j \neq k, i, j, k = 1, 2, \ldots, N$;

No mutation occurs: $j$-type particle may also remain type unchanged with probability $T^j(j)$. The two cases in which mutations occur and do not occur during the mutation phase are shown in Figure 3.
Figure 5: The two cases in which mutations occur (case 1) and do not occur (case 2) during the mutation phase.

Naturally, the branching law \( \{P^i(j)\}_{i,j=1,\ldots,N} \) of each generation remains the same. However, when mutations are considered, the branching law \( \{P^i(j)\}_{i,j=1,\ldots,N} \) of each generation will be different, i.e., the branching law matrix will change, and the iterative process is as follows.

Suppose that at time 0, there is an \( i \)-type ancestor in the system, which is the 1st generation. After a unit of time, we observe that the \( i \)-type ancestor has a \( j \)-type offspring. Then in the 2nd generation, the probability of an \( i \)-type particle produces a \( j \)-type particle when mutations are taken into account:

\[
P^i_2(j) = P^i_1(j)T^j(j) + \sum_{k \neq j} P^i_1(k)T^k(j)
\]

where the subscript denotes the generation. The first term on the right-hand side of equation (1) is the probability of no mutation occurring in the mutation stage, and the second term on the right-hand side of equation (1) is the probability of the mutation occurring. Its matrix form is:

\[
P_2 = P_1T
\]

We call the matrix \( T \) a probability matrix of mutation. Iteratively, for the \( n \)th generation, we have

\[
P_n = P_{n-1}T = PT^{n-1}
\]

Next, we use one example to show the iteration process.
Example 3.1. (The continuation of example 2.1) Here we consider the special case, the probability matrix of mutation $T = P$. The 3-type GW forest with mutation and the mutation details are shown in Figure 6.

(a) The 3-type GW forest with mutation. (b) The mutation details of the 3-type GW forest with mutation.

Figure 6: The 3-type GW forest with mutation and the mutation details

where

$$P_2 = P \ast T = P \ast T = \begin{pmatrix} 0 & 0.6250 & 0.3750 \\ 0.3846 & 0 & 0.6154 \\ 0.2727 & 0.7273 & 0 \end{pmatrix} \ast \begin{pmatrix} 0 & 0.6250 & 0.3750 \\ 0.3846 & 0 & 0.6154 \\ 0.2727 & 0.7273 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0.3427 & 0.2727 & 0.3846 \\ 0.1678 & 0.6879 & 0.1442 \\ 0.2797 & 0.1705 & 0.5498 \end{pmatrix}$$

and

$$P_3 = P_2 \ast T = P \ast T^2 = \begin{pmatrix} 0.2098 & 0.4939 & 0.2963 \\ 0.3039 & 0.2098 & 0.4863 \\ 0.2155 & 0.5747 & 0.2098 \end{pmatrix}.$$
Definition 3.2. (communicate) Let $i, j \in S$ be two states of a Markov chain with transition matrix $T$. Then state $i$ communicates with state $j$ if there exist nonnegative integer $m$ and $n$ such that $T_{j,i}^m > 0$ and $T_{i,j}^n > 0$.

Definition 3.3. (irreducible) A Markov chain with finite-state space $S$ and transition matrix $T$ is said to be irreducible if for all $i, j \in S$ are communicated.

Definition 3.4. (aperiodic) The period $d(i)$ of a state $i$ of a Markov chain is the greatest common divisor of all time steps $n$ for which $T_{i,i}^n > 0$. If the period is 1, the state is aperiodic.

If a Markov chain is irreducible, the period of the chain is the period of its single communication class. If the period of every communication class is $d = 1$, then the Markov chain is aperiodic.

Definition 3.5. (regular) A stochastic matrix $T$ is regular if some power $T^k$ contains only strictly positive entries.

Lemma 3.1. Let $T$ be the transition matrix for an irreducible, aperiodic Markov chain. Then $T$ is a regular matrix.

Lemma 3.2. If $T$ is a regular $m \times m$ transition matrix of Markov chain with $m \geq 2$, then the following statements are all true.

1. There is a stochastic matrix $\Pi$ such that $\lim_{n \to \infty} T^n = \Pi$.
2. Each row of $\Pi$ is the same probability vector $\pi$.
3. For any initial probability vector $x_0$, $\lim_{n \to \infty} x_0 T^n = \pi$.
4. The vector $\pi$ is the unique probability vector that is an eigenvector of $T$ associated with the eigenvalue 1, i.e. $\pi = \pi T$.

As described above, to prove the convergence of the iterative process we make the following statement.

- The MGWF with mutation is a finite-state Markov chain. The state-space of MGWF with mutation $S$ is the finite type space of particles $\{1, 2, \cdots, N\}$. And the iterative process $F_n = F_{n-1} T$ implies that $F_n$ depends only on $F_{n-1}$ which is meets the Markov property.

- The MGWF with mutation is an irreducible chain. Different types of particles can mutate with each other to ensure that all states of the MGWF with mutation $\{1, 2, \cdots, N\}$ are communicating.
• The MGWF with mutation is required to be an aperiodic chain.

From Lemma 3.1, the probability matrix of the mutation $T$ is regular. Therefore, from Lemma 3.2, we have that the MGWF with the mutation has a unique stationary distribution $\Pi$, s.t.

$$\lim_{n \to \infty} T^n = \Pi$$

where Each row of $\Pi$ is the same probability vector $\pi$ which is the unique probability vector that is an eigenvector of $T$ associated with the eigenvalue 1.

When the system of the MGWF with mutations reaches a stable state $\Pi$, the branching law is no longer affected by the mutation.

Example 3.2. The continuation of example 3.1 We still consider the 3-type GW forest with mutation. Where the branching law matrix is

$$P = \begin{pmatrix} 0 & 0.6250 & 0.3750 \\ 0.3846 & 0 & 0.6154 \\ 0.2727 & 0.7273 & 0 \end{pmatrix}$$

and the probability matrix of mutation $T = P$. We verify that $k = 2, T^2_{ij} > 0$ for all $i,j \in \{1,2,\cdots,N\}$, so $T$ is regular, satisfying the Lemma 3.2 condition. We set the accuracy to $1e^{-4}$ and found that when $n = 88$ the

$$P^{88} = \begin{pmatrix} 0.2500 & 0.4062 & 0.3437 \\ 0.2500 & 0.4062 & 0.3437 \\ 0.2500 & 0.4062 & 0.3437 \end{pmatrix}$$

and iterating $P^{89}, P^{90}, \cdots$ again no longer changes, i.e., the particle mutation will not affect the branching law of the MGWF.

3.3. Interpretation of diffusion process

A diffusion process, in general, starts from a $N \times N$ affinity matrix $A$, which relates $N$ different elements to each other. We interpret the matrix $A$ as a graph $G = (V,E)$, consisting of $N$ nodes $v_i \in V$, and edges $e_{ij} \in E$ that link nodes to each other, fixing the edge weights to provided affinity values with $A_{ij}$. Diffusion processes then spread the affinity values through the entire graph, based on the defined edge weights. When using a random walk on a graph to explain the diffusion process, there are three steps: (1)
initialization; (2) definition of the transition matrix; (3) definition of the diffusion process.

In this paper, we use the model of MGWF with mutation to explain the diffusion process from a new perspective. We can regard the \( N \) data points as \( N \) different particle types, the transition of the random walk between nodes corresponds to the iteration of MGWF with mutation. And the unit of time for each iteration is divided into two phases: the branching phase and the mutation phase.

Assume the initial branching law and the probability matrix of mutation are given. To ensure that the initial branching law is a row-stochastic matrix, we take \( F_0 = P = D^{-1}A \) where \( D = \text{diag}(D_{ii}) \), \( D_{ii} = \sum_{j=1}^{N} A_{ij} \), and especially the probability matrix of mutation \( T = P = D^{-1}A \). After the first unit time, the branching law of the first generation is given by

\[
F_1 = F_0T.
\]

Since the mutation probability in the mutation phase is a constant, the branching law of the second generation is derived as

\[
F_2 = F_1T = F_0T^2,
\]

Eventually, The branching law of the \( nth \) generation is

\[
F_n = F_{n-1}T = F_0T^n.
\]

It can be seen from section 3.2 that the branching law will reach a stable state, that is, even if the mutation occurs, the branching law will no longer be affected by the mutation.

The correspondence between the diffusion process and the MGWF with the mutation is shown in Table 1.

In this section, we established the relationship between the diffusion process and the MGWF with mutation and the convergence of the diffusion process is proved indirectly via the convergence of the MGWF. Next, we will propose the immigration concept for the MGWF and discuss its relationship with the Google Pagerank system, which is an extension of the diffusion process.
Table 1: The correspondence between diffusion process and MGWF with mutation

| Diffusion process                | MGWF with mutation                      |
|---------------------------------|-----------------------------------------|
| N nodes in the graph            | N types of particle                     |
| Edges in the graph              | Particle type change                    |
| An iteration of the diffusion process | A unit time of MGWF                   |
| Initialization $F_0 = P$        | Initial branching law matrix $F_0 = P$  |
| Transition matrix $P$           | The probability matrix of mutation $T = P$ |
| Update scheme $F_n = F_{n-1}P$  | Update scheme $F_n = F_{n-1}T$          |

4. The MGWF with immigration and the Google PageRank system

4.1. The MGWF with mutation and immigration

The classic MGWF has a wide range of extensions, and one of the important extensions is the GW process with immigration [37, 38, 36, 39]. The MGWF with immigration in the existing literature can be divided into states-dependent immigration and states-independent immigration. In this article, we consider states-independent immigration. The classical theory of immigration is to introduce immigrants to ensure that the entire forest is not extinct so that immigrant particles would enter the forest and can produce offspring together with the original particles, as shown in Figure 7(a). In this section, we only consider immigration particles as another option different from the original branching to produce offspring in addition to the original branching. That is, the branching-mutation mechanism will no longer work if immigrants enter the forest.
(a) The 3-type branching process with immigration.

(b) The relationship between immigration mechanism and the branching-mutation mechanism in the MGWF.

Figure 7: The immigration in branching process and MGWF with mutation

In a \(N\)-type GWF system, for every tree and every generation, there may be an immigrant of \(j\)-type particle entering with probability \(b_j\), where \(j = 1, 2, \cdots, N\) and \(\sum_{j=1}^{N} b_j = 1\). We allow migrants to enter a MGWF with mutations systems. When it is observed that an \(i\)-type particle in the 1-st generation has a child particle of \(j\)-type in the second generation, there are two possible mechanisms for this phenomenon:

(1) Generated by a branching-mutation mechanism with probability \(\alpha\), i.e. An \(i\)-type particle produces a \(j\)-type particle in consideration of mutations. Where \(\alpha \in (0, 1)\).

(2) Acted by the immigration mechanism with probability \((1 - \alpha)\). i.e. in the second generation, an immigrant particle of \(j\)-type enters directly, the branching - mutations mechanism no longer works.

Therefore, the branching law in the second generation will change as follows:

\[
P_2(j) = \alpha \left( P_1^i(j)T^j(j) + \sum_{k \neq j} P_1^i(k)T^k(j) \right) + (1 - \alpha)b_j, \quad i, j, k = 1, 2, \cdots, N
\]

The matrix form is

\[
P_2 = \alpha P_1 T + (1 - \alpha)Y
\]

where

\[
Y = \begin{pmatrix}
    b_1 & b_2 & \cdots & b_N \\
    b_1 & b_2 & \cdots & b_N \\
    \vdots & \vdots & \ddots & \vdots \\
    b_1 & b_2 & \cdots & b_N
\end{pmatrix}
\]
Figure 7(b) shows the relationship between the immigration mechanism and the branching-mutation mechanism. The left side of the figure represents the immigration mechanism, while the right side represents branching - mutation mechanism. The convergence of this process is given below.

\[ F_0 = P, T = P \]

\[ F_n = \alpha F_{n-1} T + (1 - \alpha)Y = \alpha^n P^{n+1} + (1 - \alpha)Y \sum_{k=0}^{n-1}\alpha^k P^k \]

Since \( T = P \) is regular which is explained in section 3.2 that meets the condition of Lemma 3.2, the limit \( \lim_{n \to \infty} P^{n+1} \) exists. And for \( \sum_{k=0}^{n-1}\alpha^k P^k \), since the eigenvalue of \( \alpha P \), \( \rho(\alpha P) < 1 \), the limit of \( \sum_{k=0}^{n-1}(\alpha P)^k \) exists when \( n \to \infty \). Consequently, the \( \lim_{n \to \infty} f_n \) exists.

4.2. The Google PageRank system

There are optimal iteration algorithms for the diffusion process, one of the most successful methods is the Google PageRank system. The system was originally designed to objectively rank webpages by measuring people’s interest in the relevant webpages and taking into account the potential hyperlink structure. In the Google PageRank system, in addition to walking on neighboring nodes with a probability \( \alpha \in (0, 1) \), the particle is also considered to jump to an arbitrary node with a very small probability \( (1 - \alpha) \), following the update mechanism below:

\[ F_{n+1} = \alpha F_n P + (1 - \alpha)Y \]

\( N \times N \) matrix \( Y \) is a stacking of row vectors \( y \), and \( y \) defines the probabilities of randomly jumping to the corresponding nodes.

We use the MGWF with mutation and immigration model to explain the extension of the diffusion process. For convenience, we make the following transformation:

\[ F_{n+1} = \alpha F_n P + (1 - \alpha)Y \\
= \left( \alpha F_n \ (1 - \alpha)I \right) \begin{pmatrix} P \\ Y \end{pmatrix} \]

In the above model, the branching law of the \( n \)-th generation is \( \left( \alpha F_n \ (1 - \alpha)I \right) \). Where the branching - mutation mechanism works with probability \( \alpha \) and
the branching law matrix is $F_n$. The immigration mechanism works with probability $(1 - \alpha)$ and the branching law matrix is $I$; The probability matrix of mutation is $\begin{pmatrix} P \\ Y \end{pmatrix}$ in mutation phase, where $P$, $Y$ respectively are the mutation probability of MGWF with mutation mechanism and immigration mechanism.

We can explain the diffusion process and the Google PageRank system in the same framework model through the proposed two-phases setting. The initial branching law matrix is $F_0 = (\alpha P \ (1 - \alpha)I)$ in the branching phase. And the probability matrix of mutation is $T = \begin{pmatrix} P \\ Y \end{pmatrix}$ in the mutation phase. The iteration rule is $F_{n+1} = F_nT$. When $\alpha = 1$, it’s diffusion process $F_{n+1} = F_nP$; When $\alpha \in (0, 1)$, it’s the Google PageRank system $F_{n+1} = \alpha F_nP + (1 - \alpha)Y$. As shown in Table 2.

Table 2: The same framework model of diffusion process and Google PageRank system

| The framework | $\alpha$ | The model |
|---------------|----------|-----------|
| $F_{n+1} = (\alpha F_n \ (1 - \alpha)I) \begin{pmatrix} P \\ Y \end{pmatrix}$ | $\alpha = 1$ | The diffusion process |
| $\alpha \in (0, 1)$ | | The Google PageRank system |

In summary, we proposed the two-phase MGWF setting and use it to explain the diffusion process and the Google PageRank system. This gives us a new perspective to understand the random walk. Further, we can use the MGWF framework to extend the diffusion process and speed up the iteration process as shown in the examples below.

5. Simulations

Example 5.1. The continuation of Example 3.2

We consider a diffusion process where the initial value $F_0 = P$, the transition matrix $P$, and the iteration rule $F_{n+1} = F_nP$ with $P$ given below. We explain the diffusion process by random walk and MGWF respectively.

$$P = \begin{pmatrix} 0 & 0.6250 & 0.3750 \\ 0.3846 & 0 & 0.6154 \\ 0.2727 & 0.7273 & 0 \end{pmatrix}.$$
(i) Random walk on graph

For the initial value $F_0 = P$, the probability that the particle starts at vertices 1, 2, and 3 is $(0.0, 0.625, 0.375)$, $(0.3846, 0.0, 0.6154)$ and $(0.2727, 0.7273, 0.0)$. Although [33, 3] has verified that the selection of initial value is more efficient for convergence, the reason for such selection cannot be justified explicitly in the random walk. With the transition matrix $P$, and the iteration rule $F_{n+1} = F_n P$, we use the binary norm of the matrix $\|F_n - F_{n-1}\|_2$ to measure the convergence of $F_n$ with an accuracy of $1e^{-8}$. We find that the diffusion process can reach this approximate stable state through 54 iterations. The random walk on the graph with 3 nodes is shown in Figure 8.

(ii) MGWF

In MGWF with mutation, the initial branching law is $F_0 = P$, such as the probability of the parent of 1-type producing children of 1, 2, 3-type is 0, 0.625, 0.375 respectively. From the figure above, it can be clearly observed that each unit time is divided into two phases in the $n$-th generation, where in the branching phase of the branching law is $F_n$, the probability matrix of mutation in the mutation phase is $T = P$. In order to get the branching law of $(n+1)$-th generation, we follow the result $F_{n+1} = F_n T$. It’s shown in
It is also found that the approximate stable state with an accuracy of $1e^{-8}$ is basically reached after 54 iterations through simulation.

From the above description, we find that the model of MGWF with the mutation can intuitively explain the diffusion process better. Moreover, we can make the branching law matrix be the same in the branching phase, but the probability matrix of mutation is different in the odd iteration and even iteration mutation phase. In the odd iteration, the probability matrix of mutation is $T$, in the even iteration, we change the probability matrix of mutation to $\lambda T + (1 - \lambda)I$ which is different from $T$ with $\lambda \in (0, 1)$ and $I$ is an identity matrix. Such a slight change leads to a different diffusion process:

$$F_{n+1} = \begin{cases} F_nT, & \text{if } n \text{ is odd;} \\ F_n(\lambda T + (1 - \lambda)I), & \text{if } n \text{ is even.} \end{cases}$$

Let us take $\lambda = \frac{1}{3}$ and apply the revised diffusion process to the example of $N=3$ above. We also use the two-norm of the matrix $\|F_n - F_{n-1}\|_2$ to measure the convergence of $F_n$ with an accuracy of $1e^{-8}$. We found through the simulation that the diffusion process reached an approximate stable state after 17 iterations, which was significantly more efficient than the original diffusion process.

The convergence behaviors of the original diffusion process and the revised diffusion process is shown in Figure 10.

![Figure 10: The convergence behaviors of the diffusion and the modified diffusion process](image-url)
Example 5.2. We consider a Google PageRank system with the initial value $F_0 = P$, the transition matrix $P$, the probability matrix of random jump $Y$, and the iteration rule $F_{n+1} = \alpha F_n P + (1 - \alpha)Y$. We explain the Google PageRank system by MGWF with mutation and immigration with $\alpha = 0.9$,

\[
P = \begin{pmatrix}
0 & 0.6250 & 0.3750 \\
0.3846 & 0 & 0.6154 \\
0.2727 & 0.7273 & 0
\end{pmatrix}.
\]

\[
Y = \begin{pmatrix}
0 & 0.4258 & 0.5742 \\
0.2514 & 0 & 0.7486 \\
0.4166 & 0.5834 & 0
\end{pmatrix}
\]

We use the two-phase theory in MGWF with mutation and immigration to explain the Google PageRank system. The initial branching law is $(\alpha P (1 - \alpha)I)$. The probability matrix of mutation is $(P \ Y)$ in the mutation phase. In order to get the branching law of $(n + 1)$-th generation, we follow the result $(\alpha F_n (1 - \alpha)I) (P \ Y)$. We use the norm $\|F_n - F_{n-1}\|_2$ to measure the convergence of $F_n$ with an accuracy of $1e^{-8}$. We find that the Google PageRank system can reach this approximate stable state through 42 iterations.

Whereas the random walk can’t explain the Google PageRank system, the Google PageRank system is perfectly explained by considering the immigration in the MGWF with mutation. Similar to the basic diffusion process, we propose a modified PageRank system by dividing each unit of time into two phases:

(i) As in Example 5.1, we also made some minor changes to the model of the Google PageRank system. We make the branching law matrix is the same in the branching phase, and the probability matrix of mutation different in the odd iteration and even iteration mutation phase. In the odd iteration, the probability matrix of mutation is $(P \ Y)$, in the even iteration, we change the probability matrix of mutation to $\lambda (P \ Y) + (1 - \lambda) (I \ J)$:
\[ F_{n+1} = \begin{cases} 
\alpha F_n \begin{pmatrix} (1 - \alpha)I \end{pmatrix} \begin{pmatrix} P \end{pmatrix}, & \text{if } n \text{ is odd;} \\
\alpha F_n \begin{pmatrix} (1 - \alpha)I \end{pmatrix} \left[ \lambda \begin{pmatrix} P \end{pmatrix} + (1 - \lambda) \begin{pmatrix} I \\ J \end{pmatrix} \right], & \text{if } n \text{ is even.}
\end{cases} \]

where \( \lambda \in (0, 1) \), \( I \) is an identity matrix and
\[
J = \begin{pmatrix} 0 & 0.9086 & 0.0914 \\ 0.8412 & 0 & 0.1588 \\ 0.5715 & 0.4285 & 0 \end{pmatrix}.
\]
is another probability matrix of mutation in immigration mechanism which is different from \( Y \).

Let us take \( \lambda = \frac{2}{3} \) and apply the revised Google PageRank system to the example with \( N=3 \) above. We found that the system has reached two different stable states in the 7th iteration and the 8th iteration respectively, which makes the norm \( \|F_n - F_{n-1}\|_2 \) a fixed constant. The speed of reaching stability of the revised Google PageRank was faster than that of the original Google PageRank system. We will conduct an in-depth study on the relationship and the choice between the two stable states in the future.

The convergence behaviors of the Google PageRank system and the revised Google PageRank system is shown in Figure 11.

![Figure 11: The convergence behaviors of the Google PageRank system and the revised Google PageRank system](image)

The Google PageRank system and the revised Google PageRank system.
(ii) We can make a couple of other changes to look at the branching-mutation mechanism and the immigration mechanism, whose change in mutation probability is responsible for the rate of convergence of $F_n$.

(a) (The change in the branching – mutation mechanism) In the odd and even iterations, the different probability matrices of mutation in the branching-mutation mechanism are selected respectively and the probability matrices of mutation in the immigration mechanism are the same.

$$
F_{n+1} = \begin{cases} 
(\alpha F_n (1 - \alpha)I) \left( \begin{array}{c} P \\ Y \end{array} \right), & \text{if } n \text{ is odd;} \\
(\alpha F_n (1 - \alpha)I) \left( \begin{array}{c} \lambda P + (1 - \lambda)I \\ Y \end{array} \right), & \text{if } n \text{ is even.}
\end{cases}
$$

We took $\lambda = \frac{2}{5}$ and found that the system would reach two different stable states in the 7th iteration and the 8th iteration respectively.

(b) (The change in the immigration mechanism) In the odd and even iterations, different probability matrix of immigration $Y$ is selected respectively.

$$
F_{n+1} = \begin{cases} 
(\alpha F_n (1 - \alpha)I) \left( \begin{array}{c} P \\ Y \end{array} \right), & \text{if } n \text{ is odd;} \\
(\alpha F_n (1 - \alpha)I) \left( \begin{array}{c} P \\ \lambda Y + (1 - \lambda)J \end{array} \right), & \text{if } n \text{ is even.}
\end{cases}
$$

We took $\lambda = \frac{2}{5}$ and found that the system reached two different stable states in the 11th iteration and the 12th iteration respectively.

According to the iteration results of (a) and (b) examples, it is found that changing the probability matrix of mutation in the branching-mutation mechanism is the key technique to improve the iteration speed. The convergence behaviors of (a),(b) are shown in Figure 12.
The change in the branching-mutation mechanism

The change in the immigration mechanism

Figure 12: The convergence behaviors of the change of probability matrix of mutation in the branching mechanism and in the immigration mechanism

6. Conclusions

In this paper, we have connected the two seemingly unrelated fields of diffusion process and branching process and established an explicit interpretation of the correspondence between the diffusion process and MGWF. Then we focused on MGWF with degenerated branching law, and innovatively proposed the concept of particle mutation in MGWF with degenerated branching law. Through the variant of MGWF, the diffusion process can be better interpreted, and we can observe the movement of particles from the longitudinal view more clearly. We divided each step of the diffusion process into two phases - branching phase and mutation phase, which not only interprets the mechanics more clear but also removes the existing limitation that the transition probability $P$ is dependent on the affinity matrix $A$. By extending MGWF with the concept of immigration, we further connected the MGWF model with the popular Google PageRank system, providing a new perspective to the latter as well. Both the diffusion process and the Google PageRank are well explained with the two-phases idea, as shown in Table 3.

In the future, we will focus on the following directions

- From Table 3, we found that the existing improvements of the diffusion process are concentrated in the branching phase, that is, to change the branching law by iteration. However, for the mutation phase, the mutation probability of all algorithms is constant. Therefore, in the future,
Table 3: The diffusion process and the Google PageRank are explained with the two-phases idea

| Method                        | The branching law | The probability of mutation |
|-------------------------------|-------------------|-----------------------------|
| Diffusion process             | $F_{n}$           | $P$                         |
| Google pagerank system        | $(\alpha F_{n} (1 - \alpha)I)$ | ($P\ Y^*$)                 |

we will pay more attention to changing the probability of mutation at each unit time.

- In Example 5.1 we found that the rate of convergence of the diffusion process was significantly accelerated after we made a slight change in the mutation probability. We will study how to change the probability of mutation to accelerate the rate of convergence.

- In Example 5.2 after slight changes were made to the mutation probability of the Google PageRank system, we found that although the rate of convergence of the Google PageRank system was significantly accelerated, two stable states appeared. It is of interest to study how to deal with two stable states.

References

[1] Ibe, Oliver. Markov processes for stochastic modeling. Newnes, 2013.

[2] Yang, Xingwei, Suzan Koknar-Tezel, and Longin Jan Latecki. "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval." 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009.

[3] Bai, Xiang, et al. "Learning context-sensitive shape similarity by graph transduction." IEEE Transactions on Pattern Analysis and Machine Intelligence 32.5 (2009): 861-874.

[4] Egozi, Amir, Yosi Keller, and Hugo Guterman. "Improving shape retrieval by spectral matching and meta similarity." IEEE Transactions on Image Processing 19.5 (2010): 1319-1327.
[5] Wang, Bo, and Zhuowen Tu. "Affinity learning via self-diffusion for image segmentation and clustering." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.

[6] Donoser, Michael, and Horst Bischof. "Diffusion processes for retrieval revisited." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.

[7] Li, Qilin, Wanquan Liu, and Ling Li. "Affinity learning via a diffusion process for subspace clustering." Pattern Recognition 84 (2018): 39-50.

[8] Donoser, Michael. "Replicator Graph Clustering." BMVC. 2013.

[9] Chen, Shuhan, et al. "Discriminative saliency propagation with sink points." Pattern recognition 60 (2016): 2-12.

[10] Lu, Song, Vijay Mahadevan, and Nuno Vasconcelos. "Learning optimal seeds for diffusion-based salient object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[11] Jiang, Jiayan, Bo Wang, and Zhuowen Tu. "Unsupervised metric learning by self-smoothing operator." 2011 International Conference on Computer Vision. IEEE, 2011.

[12] Li, Qilin, et al. "Semisupervised learning on graphs with an alternating diffusion process." IEEE Transactions on Neural Networks and Learning Systems (2020).

[13] Zhu, Xiaojin. Semi-supervised learning with graphs. Carnegie Mellon University, 2005.

[14] Wang, Tao, et al. "Interactive Image Segmentation Based on Label Pair Diffusion." IEEE Transactions on Industrial Informatics 17.1 (2020): 135-146.

[15] Wang, T., et al. "Label Group Diffusion for Interactive Image Segmentation," Pattern Recognit., vol. 112, pp. 107789, 2021.

[16] Guimarães Pedronette, Daniel Carlos, Lucas Pascotti Valem, and Longin Jan Latecki. "Efficient Rank-Based Diffusion Process with Assured Convergence." Journal of Imaging 7.3 (2021): 49.
[17] Ye, Jianfeng, et al. "Affinity Learning Via Self-Supervised Diffusion for Spectral Clustering." IEEE Access 9 (2020): 7170-7182.

[18] Magliani, Federico, and Andrea Prati. "LSH kNN graph for diffusion on image retrieval." Information Retrieval Journal 24.2 (2021): 114-136.

[19] Yang, Xingwei, Lakshman Prasad, and Longin Jan Latecki. "Affinity learning with diffusion on tensor product graph." IEEE transactions on pattern analysis and machine intelligence 35.1 (2012): 28-38.

[20] Wang, Bo, and Zhuowen Tu. "Sparse subspace denoising for image manifolds." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[21] Athreya, K. B., Ney, P.E., Branching process. Berlin: Springer, 1972.

[22] Ross, Sheldon M., et al. Stochastic processes. Vol. 2. New York: Wiley, 1996.

[23] Berzunza, Gabriel. "On scaling limits of multitype Galton-Watson trees with possibly infinite variance." arXiv preprint arXiv:1605.04810 (2016).

[24] Miermont, Grégory. "Invariance principles for spatial multitype Galton-Watson trees." Annales de lIHP Probabilistic Statistics. Vol. 44. No. 6. 2008.

[25] de Raphélis, Loïc. "Scaling limit of multitype Galton–Watson trees with infinitely many types." Annales de l'Institut Henri Poincaré, Probabilités et Statistiques. Vol. 53. No. 6. 2017.

[26] Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine." Computer networks and ISDN systems 30.1-7 (1998): 107-117.

[27] Duquesne, Thomas. "A limit theorem for the contour process of conditioned Galton–Watson trees." The Annals of Probability 31.2 (2003): 996-1027.

[28] Duquesne, Thomas, and Jean-Francois Le Gall. "Random trees, Lévy processes and spatial branching processes." arXiv preprint math/0509558 (2005).
[29] Chaumont, Loïc, and Thi Ngoc Anh Nguyen. "On mutations in the branching model for multitype populations." Advances in Applied Probability 50.2 (2018): 543-564.

[30] Cheek, David, and Tibor Antal. "Mutation frequencies in a birth-death branching process." The Annals of Applied Probability 28.6 (2018): 3922-3947.

[31] P. Kumar, P. Varaiya, Stochastic Systems Estimation, Identification and Adaptive Control, ser. Information and System Sciences Series. Englewood Cliffs New Jersey: Prentice Hall, 1986.

[32] Häggström, Olle. Finite Markov chains and algorithmic applications. No. 52. Cambridge University Press, 2002.

[33] Lay D C , Lay S R , Mcdonald J J. Linear algebra and its applications/ 6th ed. Wiley-Interscience, 2021.

[34] Yang, Xingwei, Lakshman Prasad, and Longin Jan Latecki. "Affinity learning with diffusion on tensor product graph." IEEE transactions on pattern analysis and machine intelligence 35.1 (2012): 28-38.

[35] Cho, Minsu, and Kyoung Mu Lee. "Authority-shift clustering: Hierarchical clustering by authority seeking on graphs." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010.

[36] Liu, Jing Ning, and Mei Zhang. "Large deviation for supercritical branching processes with immigration." Acta Mathematica Sinica, English Series 32.8 (2016): 893-900.

[37] Seneta, Eugene. "On the supercritical Galton-Watson process with immigration." Mathematical Biosciences 7.1-2 (1970): 9-14.

[38] Seneta, Eugene. "A note on the supercritical Galton-Watson process with immigration." Mathematical Biosciences 6 (1970): 305-311.

[39] Sun, Qi, and Mei Zhang. "Harmonic moments and large deviations for supercritical branching processes with immigration." Frontiers of Mathematics in China 12.5 (2017): 1201-1220.