Understanding rare events in graphs and networks

Sourav Chatterjee

Stanford University

Large deviations

- The theory of large deviations aims to study two things:

Large deviations

- The theory of large deviations aims to study two things:
  (a) the probabilities of rare events; and

Large deviations

- The theory of large deviations aims to study two things:
  (a) the probabilities of rare events; and
  (b) the conditional distributions of random variables given that some rare event has occurred (that is, what does the world look like if we know that some rare event has happend?).

Large deviations

- The theory of large deviations aims to study two things:
  (a) the probabilities of rare events; and
  (b) the conditional distributions of random variables given that some rare event has occurred (that is, what does the world look like if we know that some rare event has happend?).
- For example, (a) what is the chance that you will win the lottery, and (b) how is your life going to change if you win the lottery?

Large deviations

- The theory of large deviations aims to study two things:
  (a) the probabilities of rare events; and
  (b) the conditional distributions of random variables given that some rare event has occurred (that is, what does the world look like if we know that some rare event has happend?).
- For example, (a) what is the chance that you will win the lottery, and (b) how is your life going to change if you win the lottery?
- Often, the second question is more interesting than the first; but it is usually essential to answer the first question to be able to understand how to approach the second.

A simple example

- Toss a fair coin n times, where n is a large number.

A simple example

- Toss a fair coin n times, where n is a large number.
- Under normal circumstances, you expect to get approximately $n/2$ heads. Also, you expect to get roughly $n/4$ pairs of consecutive heads.

A simple example

- Toss a fair coin n times, where n is a large number.
- Under normal circumstances, you expect to get approximately $n/2$ heads. Also, you expect to get roughly $n/4$ pairs of consecutive heads.
- However, suppose that the following rare event occurs: you get approximately $2n/3$ heads.

A simple example

- Toss a fair coin n times, where n is a large number.
- Under normal circumstances, you expect to get approximately $n/2$ heads. Also, you expect to get roughly $n/4$ pairs of consecutive heads.
- However, suppose that the following rare event occurs: you get approximately $2n/3$ heads.
- General-purpose tools from the theory of large deviations allows us to compute that the probability of this rare event is approximately

$$e^{-n\log(2^{5/3}/3)}.$$

A simple example

- Toss a fair coin n times, where n is a large number.
- Under normal circumstances, you expect to get approximately $n/2$ heads. Also, you expect to get roughly $n/4$ pairs of consecutive heads.
- However, suppose that the following rare event occurs: you get approximately $2n/3$ heads.
- General-purpose tools from the theory of large deviations allows us to compute that the probability of this rare event is approximately

$$e^{-n \log(2^{5/3}/3)}.$$

- Moreover, we can make conclusions like the following: If this rare event has occurred, then it is highly likely that there are approximately $4n/9$ pairs of consecutive heads.

How is this estimate obtained?

- Let $X_1, \ldots, X_n$ be independent random variables, such that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$ for each i. Then the number of heads in n tosses of a fair coin has the same distribution as $S_n := X_1 + \cdots + X_n$.

How is this estimate obtained?

- Let $X_1, \ldots, X_n$ be independent random variables, such that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$ for each i. Then the number of heads in n tosses of a fair coin has the same distribution as $S_n := X_1 + \cdots + X_n$.
- For any $\theta \geq 0$,

$$\begin{aligned}
\mathbb{P}(S_n \geq 2n/3) &= \mathbb{P}(e^{\theta S_n} \geq e^{2\theta n/3}) \leq \frac{\mathbb{E}(e^{\theta S_n})}{e^{2\theta n/3}} \qquad \text{(Markov's inequality)} \\
&= \frac{\mathbb{E}(\prod_{i=1}^n e^{\theta X_i})}{e^{2\theta n/3}} = \frac{\prod_{i=1}^n \mathbb{E}(e^{\theta X_i})}{e^{2\theta n/3}} \qquad \text{(Independence)} \\
&= e^{-2\theta n/3}\left(\frac{1+e^{\theta}}{2}\right)^n.
\end{aligned}$$

How is this estimate obtained?

- Let $X_1, \ldots, X_n$ be independent random variables, such that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$ for each i. Then the number of heads in n tosses of a fair coin has the same distribution as $S_n := X_1 + \cdots + X_n$.
- For any $\theta \geq 0$,

$$\begin{aligned}
\mathbb{P}(S_n \geq 2n/3) &= \mathbb{P}(e^{\theta S_n} \geq e^{2\theta n/3}) \leq \frac{\mathbb{E}(e^{\theta S_n})}{e^{2\theta n/3}} \quad \text{(Markov's inequality)} \\
&= \frac{\mathbb{E}(\prod_{i=1}^n e^{\theta X_i})}{e^{2\theta n/3}} = \frac{\prod_{i=1}^n \mathbb{E}(e^{\theta X_i})}{e^{2\theta n/3}} \quad \text{(Independence)} \\
&= e^{-2\theta n/3}\left(\frac{1+e^\theta}{2}\right)^n.
\end{aligned}$$

- Optmizing over θ gives the desired upper bound. Lower bound involves a different idea.

Linearity

- The previous example has a built-in linearity, which allows us to explicitly compute $\mathbb{E}(e^{\theta S_n})$.

Linearity

- The previous example has a built-in linearity, which allows us to explicitly compute $\mathbb{E}(e^{\theta S_n})$.
- Generalizing this idea, classical large deviations theory possesses a collection of powerful tools to deal with linear functionals of independent random variables, or vectors, or more abstract random objects.

Linearity

- The previous example has a built-in linearity, which allows us to explicitly compute $\mathbb{E}(e^{\theta S_n})$.
- Generalizing this idea, classical large deviations theory possesses a collection of powerful tools to deal with linear functionals of independent random variables, or vectors, or more abstract random objects.
- But: No general tools for nonlinear functionals.

Linearity

- The previous example has a built-in linearity, which allows us to explicitly compute $\mathbb{E}(e^{\theta S_n})$.
- Generalizing this idea, classical large deviations theory possesses a collection of powerful tools to deal with linear functionals of independent random variables, or vectors, or more abstract random objects.
- But: No general tools for nonlinear functionals.
- A simple, natural example is given in the next slide.

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.
- Rare events in random graphs and networks are often nonlinear in nature.

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.
- Rare events in random graphs and networks are often nonlinear in nature.
- For example, consider the following simple model:

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.
- Rare events in random graphs and networks are often nonlinear in nature.
- For example, consider the following simple model:
  - There are N individuals.

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.
- Rare events in random graphs and networks are often nonlinear in nature.
- For example, consider the following simple model:
  - There are N individuals.
  - Any two are friends with probability p, not friends with probability $1 - p$. Friendships form independently.

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.
- Rare events in random graphs and networks are often nonlinear in nature.
- For example, consider the following simple model:
  - There are N individuals.
  - Any two are friends with probability p, not friends with probability $1 - p$. Friendships form independently.
  - This is known as the Erdős–Rényi $G(N, p)$ model. Not realistic, but a first step to understanding real networks.

Rare events in large networks

- Understanding real world networks is an important area of research at the present time.
- Rare events in random graphs and networks are often nonlinear in nature.
- For example, consider the following simple model:
  - There are N individuals.
  - Any two are friends with probability p, not friends with probability $1 - p$. Friendships form independently.
  - This is known as the Erdős–Rényi $G(N, p)$ model. Not realistic, but a first step to understanding real networks.
- In graph theoretic terminology, an individual is called a vertex and a friendship between two individuals is called an edge.

A rare event in the $G(N, p)$ model

- The expected number of triangles in a $G(N, p)$ random graph is $N(N-1)(N-2)p^3/6$.

A rare event in the $G(N, p)$ model

- The expected number of triangles in a $G(N, p)$ random graph is $N(N-1)(N-2)p^3/6$.
- Large deviation questions: (a) What is the probability that there are k triangles, where k is some number much bigger than the expected value?

A rare event in the $G(N, p)$ model

- The expected number of triangles in a $G(N, p)$ random graph is $N(N-1)(N-2)p^3/6$.
- Large deviation questions: (a) What is the probability that there are k triangles, where k is some number much bigger than the expected value? (b) What does the graph look like, if such a rare event occurs?

A rare event in the $G(N, p)$ model

- The expected number of triangles in a $G(N, p)$ random graph is $N(N-1)(N-2)p^3/6$.
- Large deviation questions: (a) What is the probability that there are k triangles, where k is some number much bigger than the expected value? (b) What does the graph look like, if such a rare event occurs?
- This is an example of a nonlinear problem, because the number of triangles is a nonlinear function of the adjacency matrix.

A rare event in the $G(N, p)$ model

- The expected number of triangles in a $G(N, p)$ random graph is $N(N-1)(N-2)p^3/6$.
- Large deviation questions: (a) What is the probability that there are k triangles, where k is some number much bigger than the expected value? (b) What does the graph look like, if such a rare event occurs?
- This is an example of a nonlinear problem, because the number of triangles is a nonlinear function of the adjacency matrix.
- Until even a few years ago, large deviations theory did not have the tools to answer such basic questions about networks.

Large deviations for the Erdős–Rényi graph

- The large deviation theory for the Erdős–Rényi random graph was developed in C. & Varadhan (2011).

Large deviations for the Erdős–Rényi graph

- The large deviation theory for the Erdős–Rényi random graph was developed in C. & Varadhan (2011).
- The theory brought together ideas and tools from classical large deviations and results from combinatorics and graph theory, such as Szemerédi's regularity lemma and the theory of graph limits, developed by Lovász and coauthors between 2004 and 2010.

An example

- Recall the Erdős–Rényi $G(N, p)$ model.

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

happens, where δ is a given positive constant?

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?
- Surprisingly, the large deviation theory implies that both scenarios can happen.

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?
- Surprisingly, the large deviation theory implies that both scenarios can happen.
- There exist $0 < \delta_1 < \delta_2$ so that:

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?

- Surprisingly, the large deviation theory implies that both scenarios can happen.
- There exist $0 < \delta_1 < \delta_2$ so that:
  - If $0 < \delta \leq \delta_1$ or $\delta \geq \delta_2$, then conditional on the event E, the graph behaves like $G(N, r)$ for some $r > p$.

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?

- Surprisingly, the large deviation theory implies that both scenarios can happen.
- There exist $0 < \delta_1 < \delta_2$ so that:
  - If $0 < \delta \leq \delta_1$ or $\delta \geq \delta_2$, then conditional on the event E, the graph behaves like $G(N, r)$ for some $r > p$.
  - If $\delta_1 < \delta < \delta_2$, then the conditional structure is not Erdős–Rényi.

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?

- Surprisingly, the large deviation theory implies that both scenarios can happen.
- There exist $0 < \delta_1 < \delta_2$ so that:
  - If $0 < \delta \leq \delta_1$ or $\delta \geq \delta_2$, then conditional on the event E, the graph behaves like $G(N, r)$ for some $r > p$.
  - If $\delta_1 < \delta < \delta_2$, then the conditional structure is not Erdős–Rényi.
- Existence of δ_1 and δ_2 were established in C. & Varadhan (2011).

An example

- Recall the Erdős–Rényi $G(N, p)$ model.
- Let T be the number of triangles. Let $\mathbb{E}(T)$ be the expected number of triangles.
- What is the most likely structure of the graph if the rare event

$$E = \{T \geq (1+\delta)\mathbb{E}(T)\}$$

  happens, where δ is a given positive constant?
- For instance, are all the extra triangles contained in a small subset of vertices with high connectivity amongst themselves?
- Or do they occur because the graph has an excess number of edges spread uniformly?
- Surprisingly, the large deviation theory implies that both scenarios can happen.
- There exist $0 < \delta_1 < \delta_2$ so that:
  - If $0 < \delta \leq \delta_1$ or $\delta \geq \delta_2$, then conditional on the event E, the graph behaves like $G(N, r)$ for some $r > p$.
  - If $\delta_1 < \delta < \delta_2$, then the conditional structure is not Erdős–Rényi.
- Existence of δ_1 and δ_2 were established in C. & Varadhan (2011).
- Formulas for δ_1 and δ_2 were derived by Lubetzky & Zhao (2014).

Lessons from the previous slide, in a nutshell

- If the number triangles exceeds the expected value by a little bit or by a lot, then the most likely scenario is that there is an excess number of edges, spread uniformly.

Lessons from the previous slide, in a nutshell

- If the number triangles exceeds the expected value by a little bit or by a lot, then the most likely scenario is that there is an excess number of edges, spread uniformly.
- If the exceedance belongs to a middle range, then there is a clustering of edges in a small region; the exact nature of this is still not fully understood.

Lessons from the previous slide, in a nutshell

- If the number triangles exceeds the expected value by a little bit or by a lot, then the most likely scenario is that there is an excess number of edges, spread uniformly.
- If the exceedance belongs to a middle range, then there is a clustering of edges in a small region; the exact nature of this is still not fully understood.
- There is no way that the above results could have been 'guessed'. They are deduced from mathematical formulas, and there is no intuitive explanation.

Other applications

- The large deviation theory for the Erdős–Rényi model has been extended to other, more realistic models of random graphs.

Other applications

- The large deviation theory for the Erdős–Rényi model has been extended to other, more realistic models of random graphs.
- For example, it was applied to exponential random graph models in C. & Diaconis (2013). These models are widely used in the analysis of real social networks.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.
- It applies only to dense graphs.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.
- It applies only to dense graphs.
- A graph is called dense if most vertices are connected to a sizable fraction of the other vertices.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.
- It applies only to dense graphs.
- A graph is called dense if most vertices are connected to a sizable fraction of the other vertices.
- For example, in the Erdős–Rényi model with $N = 10000$ and $p = .3$, each vertex is connected to approximately 3000 other vertices.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.
- It applies only to dense graphs.
- A graph is called dense if most vertices are connected to a sizable fraction of the other vertices.
- For example, in the Erdős–Rényi model with $N = 10000$ and $p = .3$, each vertex is connected to approximately 3000 other vertices.
- This is not true for real networks, which are usually sparse.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.
- It applies only to dense graphs.
- A graph is called dense if most vertices are connected to a sizable fraction of the other vertices.
- For example, in the Erdős–Rényi model with $N = 10000$ and $p = .3$, each vertex is connected to approximately 3000 other vertices.
- This is not true for real networks, which are usually sparse.
- Unfortunately, the graph theoretic tools used for the analysis of large deviations for random graphs are useful only in the dense setting.

An incompleteness of the theory

- The large deviation theory for random graphs, discussed in the preceding slides, has one serious shortcoming.
- It applies only to dense graphs.
- A graph is called dense if most vertices are connected to a sizable fraction of the other vertices.
- For example, in the Erdős–Rényi model with $N = 10000$ and $p = .3$, each vertex is connected to approximately 3000 other vertices.
- This is not true for real networks, which are usually sparse.
- Unfortunately, the graph theoretic tools used for the analysis of large deviations for random graphs are useful only in the dense setting.
- For example, the development of a satisfactory version of Szemerédi's regularity lemma for sparse graphs is a mathematical challenge that has remained unsolved for forty years.

The main issue

- The key problem is to enumerate the number of graphs with some given set of properties. For example, how many graphs are there with m edges and k triangles?

The main issue

- The key problem is to enumerate the number of graphs with some given set of properties. For example, how many graphs are there with m edges and k triangles?
- For dense graphs, approximate counting is possible using Szemerédi's regularity lemma.

The main issue

- The key problem is to enumerate the number of graphs with some given set of properties. For example, how many graphs are there with m edges and k triangles?
- For dense graphs, approximate counting is possible using Szemerédi's regularity lemma.
- The regularity lemma classifies the set of dense graphs on n vertices into a bounded number of 'types'. The number of types depends on the desired accuracy of approximation. Approximate counting within each type can be achieved using classical large deviation techniques.

The main issue

- The key problem is to enumerate the number of graphs with some given set of properties. For example, how many graphs are there with m edges and k triangles?
- For dense graphs, approximate counting is possible using Szemerédi's regularity lemma.
- The regularity lemma classifies the set of dense graphs on n vertices into a bounded number of 'types'. The number of types depends on the desired accuracy of approximation. Approximate counting within each type can be achieved using classical large deviation techniques.
- The regularity lemma is inapplicable in the sparse setting. We do not know how to classify sparse graphs into types.

The main issue

- The key problem is to enumerate the number of graphs with some given set of properties. For example, how many graphs are there with m edges and k triangles?
- For dense graphs, approximate counting is possible using Szemerédi's regularity lemma.
- The regularity lemma classifies the set of dense graphs on n vertices into a bounded number of 'types'. The number of types depends on the desired accuracy of approximation. Approximate counting within each type can be achieved using classical large deviation techniques.
- The regularity lemma is inapplicable in the sparse setting. We do not know how to classify sparse graphs into types.
- For example, an excess number of triangle can occur because (a) there are extra edges distributed uniformly, or (b) a small number of vertices are highly interconnected, or (c) a small number of vertices have high connectivity to the rest, or .....

A recent development

- A general theory of nonlinear large deviations was proposed in C. & Dembo (2014), that goes beyond the graph theoretic setting and bypasses the regularity lemma.

A recent development

- A general theory of nonlinear large deviations was proposed in C. & Dembo (2014), that goes beyond the graph theoretic setting and bypasses the regularity lemma.
- Using this theory, Lubetzky & Zhao (2014) showed that if T is the number of triangles in $G(N,p)$, then as $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}$,

$$\begin{aligned}&\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))\\ &= \exp\left(-(1+o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\} N^2 p^2 \log\frac{1}{p}\right).\end{aligned}$$

A recent development

- A general theory of nonlinear large deviations was proposed in C. & Dembo (2014), that goes beyond the graph theoretic setting and bypasses the regularity lemma.
- Using this theory, Lubetzky & Zhao (2014) showed that if T is the number of triangles in $G(N,p)$, then as $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}$,

$$\begin{aligned}&\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))\\ &= \exp\left(-(1+o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\}N^2p^2\log\frac{1}{p}\right).\end{aligned}$$

- This gives an almost complete solution to a very old unsolved problem in the literature on random graphs, improving on contributions from many authors.

A recent development

- A general theory of nonlinear large deviations was proposed in C. & Dembo (2014), that goes beyond the graph theoretic setting and bypasses the regularity lemma.
- Using this theory, Lubetzky & Zhao (2014) showed that if T is the number of triangles in $G(N,p)$, then as $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}$,

$$\begin{aligned}&\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))\\ &= \exp\left(-(1+o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\} N^2 p^2 \log\frac{1}{p}\right).\end{aligned}$$

- This gives an almost complete solution to a very old unsolved problem in the literature on random graphs, improving on contributions from many authors.
- Vastly generalized in Bhattacharya, Ganguly, Lubetzky & Zhao (2015), yielding striking new formulas and breakthroughs.

Future directions and challenges

- Main problem: we do not yet understand the nature of sparse graphs. We understand some sparse graph structures, but not all possible structures in totality.

Future directions and challenges

- Main problem: we do not yet understand the nature of sparse graphs. We understand some sparse graph structures, but not all possible structures in totality.
- The problem can be solved by devising a suitable version of Szemerédi's regularity lemma for sparse graphs. None existing, as of now.

Future directions and challenges

- Main problem: we do not yet understand the nature of sparse graphs. We understand some sparse graph structures, but not all possible structures in totality.
- The problem can be solved by devising a suitable version of Szemerédi's regularity lemma for sparse graphs. None existing, as of now.
- Instead, some progress has been made using the newly developed tools of nonlinear large deviations.

Future directions and challenges

- Main problem: we do not yet understand the nature of sparse graphs. We understand some sparse graph structures, but not all possible structures in totality.
- The problem can be solved by devising a suitable version of Szemerédi's regularity lemma for sparse graphs. None existing, as of now.
- Instead, some progress has been made using the newly developed tools of nonlinear large deviations.
- However, the theory is still in a rudimentary form. Needs substantial improvements.

Future directions and challenges

- Main problem: we do not yet understand the nature of sparse graphs. We understand some sparse graph structures, but not all possible structures in totality.
- The problem can be solved by devising a suitable version of Szemerédi's regularity lemma for sparse graphs. None existing, as of now.
- Instead, some progress has been made using the newly developed tools of nonlinear large deviations.
- However, the theory is still in a rudimentary form. Needs substantial improvements.
- Lastly, from an applied perspective: there is a gap between the networks that appear in the real world, versus the networks whose large deviation properties can be theoretically analyzed. This gap needs to be bridged.

Future directions and challenges

- Main problem: we do not yet understand the nature of sparse graphs. We understand some sparse graph structures, but not all possible structures in totality.
- The problem can be solved by devising a suitable version of Szemerédi's regularity lemma for sparse graphs. None existing, as of now.
- Instead, some progress has been made using the newly developed tools of nonlinear large deviations.
- However, the theory is still in a rudimentary form. Needs substantial improvements.
- Lastly, from an applied perspective: there is a gap between the networks that appear in the real world, versus the networks whose large deviation properties can be theoretically analyzed. This gap needs to be bridged.
- More details about everything in this talk are in my recent monograph on this topic.