Properties of Random Graphs – Subgraph Containment

Lars Hupel

March 17, 2025

Abstract

Random graphs are graphs with a fixed number of vertices, where each edge is present with a fixed probability. We are interested in the probability that a random graph contains a certain pattern, for example a cycle or a clique. A very high edge probability gives rise to perhaps too many edges (which degrades performance for many algorithms), whereas a low edge probability might result in a disconnected graph. We prove a theorem about a threshold probability such that a higher edge probability will asymptotically almost surely produce a random graph with the desired subgraph.

Contents

1 Introduction

Random graphs have been introduced by Erdős and Rényi in [2]. They describe a probability space where, for a fixed number of vertices, each possible edge is present with a certain probability independent from other edges, but with the same probability for each edge. They study what properties emerge when increasing the number of vertices, or as they call it, "the evolution of such a random graph". The theorem which we will prove here is a slightly different version from that in the first section of that paper.

Here, we are interested in the probability that a random graph contains a certain pattern, for example a cycle or a clique. A very high edge probability gives rise to perhaps too many edges, which is usually undesired since it degrades the performance of many algorithms, whereas a low edge probability might result in a disconnected graph. The central theorem determines a threshold probability such that a higher edge probability will asymptotically almost surely produce a random graph with the desired subgraph.

The proof is outlined in [1, § 11.4] and [3, § 3]. The work is based on the comprehensive formalization of probability theory in Isabelle/HOL and on a previous definition of graphs in a work by Noschinski [4]. There, Noschinski formalized the proof that graphs with arbitrarily large girth and chromatic number exist. While the proof in this paper uses a different approach, the definition of a probability space on edges turned out to be quite useful.

2 Miscellaneous and contributed lemmas

theory *Ugraph-Misc*
imports
  HOL−Probability.Probability
  Girth-Chromatic.Girth-Chromatic-Misc
begin

lemma *sum-square*:
  fixes $a :: 'i \Rightarrow 'a :: \{monoid\text{-}mult,\ semiring\text{-}0\}$
  shows $(\sum i \in I.\ a\ i)\hat{\ }2 = (\sum i \in I.\ \sum j \in I.\ a\ i * a\ j)$
  ⟨*proof*⟩

lemma *sum-split*:
  $finite\ I \Longrightarrow$
  $(\sum i \in I.\ if\ p\ i\ then\ f\ i\ else\ g\ i) = (\sum i \mid i \in I \wedge p\ i.\ f\ i) + (\sum i \mid i \in I \wedge \neg p\ i.\ g\ i)$
  ⟨*proof*⟩

lemma *sum-split2*:
  assumes *finite I*
  shows $(\sum i \mid i \in I \wedge P\ i.\ if\ Q\ i\ then\ f\ i\ else\ g\ i) = (\sum i \mid i \in I \wedge P\ i \wedge Q\ i.\ f\ i) + (\sum i \mid i \in I \wedge P\ i \wedge \neg Q\ i.\ g\ i)$

⟨*proof*⟩

lemma *sum-upper*:
  fixes f :: $'i \Rightarrow 'a$ :: *ordered-comm-monoid-add*
  assumes *finite* I $\bigwedge i.\ i \in I \Longrightarrow 0 \leq f\ i$
  shows $(\sum i \mid i \in I \wedge P\ i.\ f\ i) \leq sum\ f\ I$
⟨*proof*⟩

lemma *sum-lower*:
  fixes f :: $'i \Rightarrow 'a$ :: *ordered-comm-monoid-add*
  assumes *finite* I $i \in I$ $\bigwedge i.\ i \in I \Longrightarrow 0 \leq f\ i$ $x < f\ i$
  shows $x < sum\ f\ I$
⟨*proof*⟩

lemma *sum-lower-or-eq*:
  fixes f :: $'i \Rightarrow 'a$ :: *ordered-comm-monoid-add*
  assumes *finite* I $i \in I$ $\bigwedge i.\ i \in I \Longrightarrow 0 \leq f\ i$ $x \leq f\ i$
  shows $x \leq sum\ f\ I$
⟨*proof*⟩

lemma *sum-left-div-distrib*:
  fixes f :: $'i \Rightarrow real$
  shows $(\sum i \in I.\ f\ i\ /\ x) = sum\ f\ I\ /\ x$
⟨*proof*⟩

lemma *powr-mono3*:
  fixes x::*real*
  assumes $0 < x$ $x < 1$ $b \leq a$
  shows $x\ powr\ a \leq x\ powr\ b$
⟨*proof*⟩

lemma *card-union*: *finite* $A \Longrightarrow$ *finite* $B \Longrightarrow card\ (A \cup B) = card\ A + card\ B - card\ (A \cap B)$
⟨*proof*⟩

lemma *card-1-element*:
  assumes $card\ E = 1$
  shows $\exists\, a.\ E = \{a\}$
⟨*proof*⟩

lemma *card-2-elements*:
  assumes $card\ E = 2$
  shows $\exists\, a\ b.\ E = \{a,\ b\} \wedge a \neq b$
⟨*proof*⟩

lemma *bij-lift*:
  assumes *bij-betw* f A B
  shows *bij-betw* $(\lambda e.\ f\ `\ e)\ (Pow\ A)\ (Pow\ B)$
⟨*proof*⟩

lemma *card-inj-subs*: *inj-on* f $A \Longrightarrow B \subseteq A \Longrightarrow$ *card* $(f$ ‘ $B)$ = *card* B
⟨*proof*⟩

lemma *image-comp-cong*: $(\bigwedge a.\ a \in A \Longrightarrow f\ a = f\ (g\ a)) \Longrightarrow f$ ‘ $A = f$ ‘ $(g$ ‘ $A)$
⟨*proof*⟩

abbreviation *less-fun* :: (*nat* ⇒ *real*) ⇒ (*nat* ⇒ *real*) ⇒ *bool* (**infix** ‹≪› *50*)
where
$f \ll g \equiv (\lambda n.\ f\ n\ /\ g\ n) \longrightarrow 0$

context
fixes f :: *nat* ⇒ *real*
begin

lemma *LIMSEQ-power-zero*: $f \longrightarrow 0 \Longrightarrow 0 < n \Longrightarrow (\lambda x.\ f\ x\ \hat{}\ n :: real) \longrightarrow 0$
⟨*proof*⟩

lemma *LIMSEQ-cong*:
assumes $f \longrightarrow x$ $\forall^\infty n.\ f\ n = g\ n$
shows $g \longrightarrow x$
⟨*proof*⟩
print-statement *Lim-transform-eventually*

lemma *LIMSEQ-le-zero*:
assumes $g \longrightarrow 0$ $\forall^\infty n.\ 0 \leq f\ n$ $\forall^\infty n.\ f\ n \leq g\ n$
shows $f \longrightarrow 0$
⟨*proof*⟩

lemma *LIMSEQ-const-mult*:
assumes $f \longrightarrow a$
shows $(\lambda x.\ c * f\ x) \longrightarrow c * a$
⟨*proof*⟩

lemma *LIMSEQ-const-div*:
assumes $f \longrightarrow a$ $c \neq 0$
shows $(\lambda x.\ f\ x\ /\ c) \longrightarrow a\ /\ c$
⟨*proof*⟩

end

lemma *quot-bounds*:
fixes x :: $'a$ :: *linordered-field*
assumes $x \leq x'$ $y' \leq y$ $0 < y$ $0 \leq x$ $0 < y'$
shows $x\ /\ y \leq x'\ /\ y'$
⟨*proof*⟩

lemma *less-fun-bounds*:
assumes $f' \ll g'$ $\forall^\infty n.\ f\ n \le f'\ n$ $\forall^\infty n.\ g'\ n \le g\ n$ $\forall^\infty n.\ 0 \le f\ n$ $\forall^\infty n.\ 0 < g\ n$ $\forall^\infty n.\ 0 < g'\ n$
shows $f \ll g$
⟨*proof*⟩

lemma *less-fun-const-quot*:
assumes $f \ll g$ $c \neq 0$
shows $(\lambda n.\ b * f\ n) \ll (\lambda n.\ c * g\ n)$
⟨*proof*⟩

lemma *partition-set-of-intersecting-sets-by-card*:
assumes *finite A*
shows $\{B.\ A \cap B \neq \{\}\} = (\bigcup n \in \{1..card\ A\}.\ \{B.\ card\ (A \cap B) = n\})$
⟨*proof*⟩

lemma *card-set-of-intersecting-sets-by-card*:
assumes $A \subseteq I$ *finite I* $k \le n$ $n \le card\ I$ $k \le card\ A$
shows $card\ \{B.\ B \subseteq I \wedge card\ B = n \wedge card\ (A \cap B) = k\} = (card\ A\ choose\ k) * ((card\ I - card\ A)\ choose\ (n - k))$
⟨*proof*⟩

lemma *card-dep-pair-set*:
assumes *finite A* $\bigwedge a.\ a \subseteq A \Longrightarrow finite\ (f\ a)$
shows $card\ \{(a, b).\ a \subseteq A \wedge card\ a = n \wedge b \subseteq f\ a \wedge card\ b = g\ a\} = (\sum a \mid a \subseteq A \wedge card\ a = n.\ card\ (f\ a)\ choose\ g\ a)$ (**is** $card\ ?S = ?C$)
⟨*proof*⟩

lemma *prod-cancel-nat*:
— Contributed by Manuel Eberl
fixes $f::'a \Rightarrow nat$
assumes $B \subseteq A$ **and** *finite A* **and** $\forall x \in B.\ f\ x \neq 0$
shows $prod\ f\ A\ /\ prod\ f\ B = prod\ f\ (A - B)$ (**is** $?A\ /\ ?B = ?C$)
⟨*proof*⟩

lemma *prod-id-cancel-nat*:
— Contributed by Manuel Eberl
fixes $A::nat\ set$
assumes $B \subseteq A$ **and** *finite A* **and** $0 \notin B$
shows $\prod A\ /\ \prod B = \prod (A - B)$
⟨*proof*⟩

lemma (**in** *prob-space*) *integrable-squareD*:
— Contributed by Johannes Hölzl
fixes $X :: - \Rightarrow real$
assumes $integrable\ M\ (\lambda x.\ (X\ x)\hat{}2)$ $X \in borel\text{-}measurable\ M$
shows *integrable M X*
⟨*proof*⟩

end
theory *Prob-Lemmas*
imports
  HOL−Probability.Probability
  Girth-Chromatic.Girth-Chromatic
  Ugraph-Misc
begin

3 Lemmas about probabilities

In this section, auxiliary lemmas for computing bounds on expectation and probabilites of random variables are set up.

3.1 Indicator variables and valid probability values

abbreviation *rind* :: *′a set* ⇒ *′a* ⇒ *real* **where**
rind ≡ *indicator*

lemma *product-indicator*:
  rind A x ∗ *rind B x* = *rind* (*A* ∩ *B*) *x*
⟨*proof*⟩

We call a real number 'valid' iff it is in the range 0 to 1, inclusively, and additionally 'nonzero' iff it is neither 0 nor 1.

abbreviation *valid-prob* (*p* :: *real*) ≡ *0* ≤ *p* ∧ *p* ≤ *1*
abbreviation *nonzero-prob* (*p* :: *real*) ≡ *0* < *p* ∧ *p* < *1*

A function *′a* ⇒ *real* is a 'valid probability function' iff each value in the image is valid, and similarly for 'nonzero'.

abbreviation *valid-prob-fun f* ≡ (∀ *n*. *valid-prob* (*f n*))
abbreviation *nonzero-prob-fun f* ≡ (∀ *n*. *nonzero-prob* (*f n*))

lemma *nonzero-fun-is-valid-fun*: *nonzero-prob-fun f* ⟹ *valid-prob-fun f*
⟨*proof*⟩

3.2 Expectation and variance

context *prob-space*
begin

Note that there is already a notion of independent sets (see *indep-set*), but we use the following – simpler – definition:

definition *indep A B* ⟷ *prob* (*A* ∩ *B*) = *prob A* ∗ *prob B*

The probability of an indicator variable is equal to its expectation:

lemma *expectation-indicator*:
  A ∈ *events* ⟹ *expectation* (*rind A*) = *prob A*

⟨*proof*⟩

For a non-negative random variable X, the Markov inequality gives the following upper bound:

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}$$

lemma *markov-inequality*:
  assumes $\bigwedge a.\ 0 \leq X\ a$ **and** *integrable M X* $0 < t$
  shows *prob* $\{a \in$ *space M*$.\ t \leq X\ a\} \leq$ *expectation X* / *t*
⟨*proof*⟩

$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$

lemma *variance-expectation*:
  fixes *X* :: *'a* ⇒ *real*
  assumes *integrable M* (λ*x*. (*X x*)^*2*) **and** *X* ∈ *borel-measurable M*
  shows
    integrable M (λ*x*. (*X x* − *expectation X*)^*2*) (**is** *?integrable*)
    variance X = *expectation* (λ*x*. (*X x*)^*2*) − (*expectation X*)^*2* (**is** *?variance*)
⟨*proof*⟩

A corollary from the Markov inequality is Chebyshev's inequality, which gives an upper bound for the deviation of a random variable from its expectation:

$$\Pr[|Y - \mathrm{E}[Y]| \geq s] \leq \frac{\mathrm{Var}[X]}{a^2}$$

lemma *chebyshev-inequality*:
  fixes *Y* :: *'a* ⇒ *real*
  assumes *Y-int*: *integrable M* (λ*y*. (*Y y*)^*2*)
  assumes *Y-borel*: *Y* ∈ *borel-measurable M*
  fixes *s* :: *real*
  assumes *s-pos*: *0* < *s*
  shows *prob* {*a* ∈ *space M*. *s* ≤ |*Y a* − *expectation Y*|} ≤ *variance Y* / *s*^*2*
⟨*proof*⟩

Hence, we can derive an upper bound for the probability that a random variable is 0.

corollary *chebyshev-prob-zero*:
  fixes *Y* :: *'a* ⇒ *real*
  assumes *Y-int*: *integrable M* (λ*y*. (*Y y*)^*2*)
  assumes *Y-borel*: *Y* ∈ *borel-measurable M*
  assumes *μ-pos*: *expectation Y* > *0*
  shows *prob* {*a* ∈ *space M*. *Y a* = *0*} ≤ *expectation* (λ*y*. (*Y y*)^*2*) / (*expectation Y*)^*2* − *1*
⟨*proof*⟩

end

3.3 Sets of indicator variables

This section introduces some inequalities about expectation and other values related to the sum of a set of random indicators.

locale *prob-space-with-indicators* = *prob-space* +
  fixes I :: $'i$ *set*
  assumes *finite-I*: *finite* I

  fixes A :: $'i \Rightarrow 'a$ *set*
  assumes A: A ‘ $I \subseteq$ *events*

  assumes *prob-non-zero*: $\exists\, i \in I.\ 0 < prob\ (A\ i)$
begin

We call the underlying sets $A\ i$ for each $i \in I$, and the corresponding indicator variables $X\ i$. The sum is denoted by Y, and its expectation by μ.

definition $X\ i = rind\ (A\ i)$
definition $Y\ x = (\sum i \in I.\ X\ i\ x)$

definition $\mu = expectation\ Y$

In the lecture notes, the following two relations are called $\sim$ and $\nsim$, respectively. Note that they are not the opposite of each other.

abbreviation *ineq-indep* :: $'i \Rightarrow 'i \Rightarrow bool$ **where**
ineq-indep $i\ j \equiv (i \neq j \wedge indep\ (A\ i)\ (A\ j))$

abbreviation *ineq-dep* :: $'i \Rightarrow 'i \Rightarrow bool$ **where**
ineq-dep $i\ j \equiv (i \neq j \wedge \neg indep\ (A\ i)\ (A\ j))$

definition $\Delta_a = (\sum i \in I.\ \sum j \mid j \in I \wedge i \neq j.\ prob\ (A\ i \cap A\ j))$
definition $\Delta_d = (\sum i \in I.\ \sum j \mid j \in I \wedge ineq\text{-}dep\ i\ j.\ prob\ (A\ i \cap A\ j))$

lemma Δ-*zero*:
  assumes $\bigwedge i\ j.\ i \in I \Longrightarrow j \in I \Longrightarrow i \neq j \Longrightarrow indep\ (A\ i)\ (A\ j)$
  shows $\Delta_d = 0$
⟨*proof*⟩

lemma *A-events*[*measurable*]: $i \in I \Longrightarrow A\ i \in events$
⟨*proof*⟩

lemma *expectation-X-Y*: $\mu = (\sum i{\in}I.\ expectation\ (X\ i))$
⟨*proof*⟩

lemma *expectation-X-non-zero*: $\exists\, i \in I.\ 0 < expectation\ (X\ i)$
⟨*proof*⟩

corollary μ-*non-zero*[*simp*]: $0 < \mu$

⟨*proof*⟩

lemma Δ_d*-nonneg*: *0* ≤ Δ_d
⟨*proof*⟩

corollary *μ-sq-non-zero*[*simp*]: *0* < *μ*^*2*
⟨*proof*⟩

lemma *Y-square-unfold*: (*λx.* (*Y x*)^*2*) = (*λx.* ∑ *i* ∈ *I.* ∑ *j* ∈ *I. rind* (*A i* ∩ *A j*) *x*)
⟨*proof*⟩

lemma *integrable-Y-sq*[*simp*]: *integrable M* (*λy.* (*Y y*)^*2*)
⟨*proof*⟩

lemma *measurable-Y*[*measurable*]: *Y* ∈ *borel-measurable M*
⟨*proof*⟩

lemma *expectation-Y-Δ*: *expectation* (*λx.* (*Y x*)^*2*) = *μ* + Δ_a
⟨*proof*⟩

lemma Δ*-expectation-X*: Δ_a ≤ *μ*^*2* + Δ_d
⟨*proof*⟩

lemma *prob-μ-*Δ_a: *prob* {*a* ∈ *space M. Y a* = *0*} ≤ *1* / *μ* + Δ_a / *μ*^*2* − *1*
⟨*proof*⟩

lemma *prob-μ-*Δ_d: *prob* {*a* ∈ *space M. Y a* = *0*} ≤ *1*/*μ* + Δ_d/*μ*^*2*
⟨*proof*⟩

end

end

4 Lemmas about undirected graphs

theory *Ugraph-Lemmas*
imports
Prob-Lemmas
Girth-Chromatic.Girth-Chromatic
begin

The complete graph is a graph where all possible edges are present. It is wellformed by definition.

definition *complete* :: *nat set* ⇒ *ugraph* **where**
complete V = (*V*, *all-edges V*)

lemma *complete-wellformed*: *uwellformed* (*complete V*)
⟨*proof*⟩

If the set of vertices is finite, the set of edges in the complete graph is finite.

lemma *all-edges-finite*: $\mathit{finite}\ V \Longrightarrow \mathit{finite}\ (\mathit{all\text{-}edges}\ V)$
⟨*proof*⟩

corollary *complete-finite-edges*: $\mathit{finite}\ V \Longrightarrow \mathit{finite}\ (\mathit{uedges}\ (\mathit{complete}\ V))$
⟨*proof*⟩

The sets of possible edges of disjoint sets of vertices are disjoint.

lemma *all-edges-disjoint*: $S \cap T = \{\} \Longrightarrow \mathit{all\text{-}edges}\ S \cap \mathit{all\text{-}edges}\ T = \{\}$
⟨*proof*⟩

A graph is called 'finite' if its set of edges and its set of vertices are finite.

definition $\mathit{finite\text{-}graph}\ G \equiv \mathit{finite}\ (\mathit{uverts}\ G) \wedge \mathit{finite}\ (\mathit{uedges}\ G)$

The complete graph is finite.

corollary *complete-finite*: $\mathit{finite}\ V \Longrightarrow \mathit{finite\text{-}graph}\ (\mathit{complete}\ V)$
⟨*proof*⟩

A graph is called 'nonempty' if it contains at least one vertex and at least one edge.

definition $\mathit{nonempty\text{-}graph}\ G \equiv \mathit{uverts}\ G \neq \{\} \wedge \mathit{uedges}\ G \neq \{\}$

A random graph is both wellformed and finite.

lemma (**in** *edge-space*) *wellformed-and-finite*:
assumes $E \in \mathit{Pow}\ S\text{-}\mathit{edges}$
shows $\mathit{finite\text{-}graph}\ (\mathit{edge\text{-}ugraph}\ E)\ \mathit{uwellformed}\ (\mathit{edge\text{-}ugraph}\ E)$
⟨*proof*⟩

The probability for a random graph to have e edges is p^e.

lemma (**in** *edge-space*) *cylinder-empty-prob*:
$A \subseteq S\text{-}\mathit{edges} \Longrightarrow \mathit{prob}\ (\mathit{cylinder}\ S\text{-}\mathit{edges}\ A\ \{\}) = p\ \widehat{}\ (\mathit{card}\ A)$
⟨*proof*⟩

4.1 Subgraphs

definition $\mathit{subgraph} :: \mathit{ugraph} \Rightarrow \mathit{ugraph} \Rightarrow \mathit{bool}$ **where**
$\mathit{subgraph}\ G'\ G \equiv \mathit{uverts}\ G' \subseteq \mathit{uverts}\ G \wedge \mathit{uedges}\ G' \subseteq \mathit{uedges}\ G$

lemma *subgraph-refl*: $\mathit{subgraph}\ G\ G$
⟨*proof*⟩

lemma *subgraph-trans*: $\mathit{subgraph}\ G''\ G' \Longrightarrow \mathit{subgraph}\ G'\ G \Longrightarrow \mathit{subgraph}\ G''\ G$
⟨*proof*⟩

lemma *subgraph-antisym*: $\mathit{subgraph}\ G\ G' \Longrightarrow \mathit{subgraph}\ G'\ G \Longrightarrow G = G'$
⟨*proof*⟩

lemma *subgraph-complete*:
assumes *uwellformed G*
shows *subgraph G* (*complete* (*uverts G*))
⟨*proof*⟩

corollary *wellformed-all-edges*: *uwellformed G* $\Longrightarrow$ *uedges G* $\subseteq$ *all-edges* (*uverts G*)
⟨*proof*⟩

corollary *max-edges-graph*:
assumes *uwellformed G finite* (*uverts G*)
shows *card* (*uedges G*) $\leq$ (*card* (*uverts G*))^*2*
⟨*proof*⟩

lemma *subgraph-finite*: ⟦ *finite-graph G*; *subgraph G′ G* ⟧ $\Longrightarrow$ *finite-graph G′*
⟨*proof*⟩

corollary *wellformed-finite*:
assumes *finite* (*uverts G*) **and** *uwellformed G*
shows *finite-graph G*
⟨*proof*⟩

definition *subgraphs* :: *ugraph* $\Rightarrow$ *ugraph set* **where**
subgraphs G = {*G′*. *subgraph G′ G*}

definition *nonempty-subgraphs* :: *ugraph* $\Rightarrow$ *ugraph set* **where**
nonempty-subgraphs G = {*G′*. *uwellformed G′* $\wedge$ *subgraph G′ G* $\wedge$ *nonempty-graph G′*}

lemma *subgraphs-finite*:
assumes *finite-graph G*
shows *finite* (*subgraphs G*)
⟨*proof*⟩

corollary *nonempty-subgraphs-finite*: *finite-graph G* $\Longrightarrow$ *finite* (*nonempty-subgraphs G*)
⟨*proof*⟩

4.2 Induced subgraphs

definition *induced-subgraph* :: *uvert set* $\Rightarrow$ *ugraph* $\Rightarrow$ *ugraph* **where**
induced-subgraph V G = (*V*, *uedges G* $\cap$ *all-edges V*)

lemma *induced-is-subgraph*:
V $\subseteq$ *uverts G* $\Longrightarrow$ *subgraph* (*induced-subgraph V G*) *G*
V $\subseteq$ *uverts G* $\Longrightarrow$ *subgraph* (*induced-subgraph V G*) (*complete V*)
⟨*proof*⟩

lemma *induced-wellformed*: *uwellformed G* $\Longrightarrow$ *V* $\subseteq$ *uverts G* $\Longrightarrow$ *uwellformed*

(*induced-subgraph V G*)
⟨*proof*⟩

lemma *subgraph-union-induced*:
  assumes *uverts* $H_1 \subseteq S$ **and** *uverts* $H_2 \subseteq T$
  assumes *uwellformed* H_1 **and** *uwellformed* H_2
  shows *subgraph* H_1 (*induced-subgraph S G*) ∧ *subgraph* H_2 (*induced-subgraph T G*) ⟷
    subgraph (*uverts* H_1 ∪ *uverts* H_2, *uedges* H_1 ∪ *uedges* H_2) (*induced-subgraph* (*S* ∪ *T*) *G*)
⟨*proof*⟩

lemma (**in** *edge-space*) *induced-subgraph-prob*:
  assumes *uverts* $H \subseteq V$ **and** *uwellformed H* **and** $V \subseteq$ *S-verts*
  shows *prob* {*es* ∈ *space P*. *subgraph H* (*induced-subgraph V* (*edge-ugraph es*))} = *p* ^ *card* (*uedges H*) (**is** *prob ?A* = -)
⟨*proof*⟩

4.3 Graph isomorphism

We define graph isomorphism slightly different than in the literature. The usual definition is that two graphs are isomorphic iff there exists a bijection between the vertex sets which preserves the adjacency. However, this complicates many proofs.

Instead, we define the intuitive mapping operation on graphs. An isomorphism between two graphs arises if there is a suitable mapping function from the first to the second graph. Later, we show that this operation can be inverted.

fun *map-ugraph* :: (*nat* ⇒ *nat*) ⇒ *ugraph* ⇒ *ugraph* **where**
map-ugraph f (*V*, *E*) = (*f* ‘ *V*, (λ*e*. *f* ‘ *e*) ‘ *E*)

definition *isomorphism* :: *ugraph* ⇒ *ugraph* ⇒ (*nat* ⇒ *nat*) ⇒ *bool* **where**
isomorphism G_1 G_2 *f* ≡ *bij-betw f* (*uverts* G_1) (*uverts* G_2) ∧ G_2 = *map-ugraph f* G_1

abbreviation *isomorphic* :: *ugraph* ⇒ *ugraph* ⇒ *bool* (‹- ≃ -›) **where**
$G_1 \simeq G_2$ ≡ *uwellformed* G_1 ∧ *uwellformed* G_2 ∧ (∃*f*. *isomorphism* G_1 G_2 *f*)

lemma *map-ugraph-id*: *map-ugraph id* = *id*
⟨*proof*⟩

lemma *map-ugraph-trans*: *map-ugraph* (*g* ∘ *f*) = (*map-ugraph g*) ∘ (*map-ugraph f*)
  ⟨*proof*⟩

lemma *map-ugraph-wellformed*:
  assumes *uwellformed G* **and** *inj-on f* (*uverts G*)
  shows *uwellformed* (*map-ugraph f G*)

⟨*proof*⟩

lemma *map-ugraph-finite*: *finite-graph* $G \Longrightarrow$ *finite-graph* (*map-ugraph f G*)
⟨*proof*⟩

lemma *map-ugraph-preserves-sub*:
 assumes *subgraph* G_1 G_2
 shows *subgraph* (*map-ugraph f* G_1) (*map-ugraph f* G_2)
⟨*proof*⟩

lemma *isomorphic-refl*: *uwellformed* $G \Longrightarrow G \simeq G$
⟨*proof*⟩

lemma *isomorphic-trans*:
 assumes $G_1 \simeq G_2$ **and** $G_2 \simeq G_3$
 shows $G_1 \simeq G_3$
⟨*proof*⟩

lemma *isomorphic-sym*:
 assumes $G_1 \simeq G_2$
 shows $G_2 \simeq G_1$
⟨*proof*⟩

lemma *isomorphic-cards*:
 assumes $G_1 \simeq G_2$
 shows
  card (*uverts* G_1) = *card* (*uverts* G_2) (**is** *?V*)
  card (*uedges* G_1) = *card* (*uedges* G_2) (**is** *?E*)
⟨*proof*⟩

4.4 Isomorphic subgraphs

The somewhat sloppy term 'isomorphic subgraph' denotes a subgraph which is isomorphic to a fixed other graph. For example, saying that a graph contains a triangle usually means that it contains *any* triangle, not the specific triangle with the nodes 1, 2 and 3. Hence, such a graph would have a triangle as an isomorphic subgraph.

definition *subgraph-isomorphic* :: *ugraph* ⇒ *ugraph* ⇒ *bool* (‹- ⊑ -›) **where**
$G' \sqsubseteq G \equiv$ *uwellformed* $G \wedge (\exists\, G''.\ G' \simeq G'' \wedge$ *subgraph* $G''\ G)$

lemma *subgraph-is-subgraph-isomorphic*: ⟦ *uwellformed* G'; *uwellformed* G; *subgraph* G' G ⟧ $\Longrightarrow G' \sqsubseteq G$
⟨*proof*⟩

lemma *isomorphic-is-subgraph-isomorphic*: $G_1 \simeq G_2 \Longrightarrow G_1 \sqsubseteq G_2$
⟨*proof*⟩

lemma *subgraph-isomorphic-refl*: *uwellformed* $G \Longrightarrow G \sqsubseteq G$

⟨*proof*⟩

lemma *subgraph-isomorphic-pre-iso-closed*:
assumes $G_1 \simeq G_2$ **and** $G_2 \sqsubseteq G_3$
shows $G_1 \sqsubseteq G_3$
⟨*proof*⟩

lemma *subgraph-isomorphic-pre-subgraph-closed*:
assumes *uwellformed* G_1 **and** *subgraph* G_1 G_2 **and** $G_2 \sqsubseteq G_3$
shows $G_1 \sqsubseteq G_3$
⟨*proof*⟩

lemmas *subgraph-isomorphic-pre-closed* = *subgraph-isomorphic-pre-subgraph-closed* *subgraph-isomorphic-pre-iso-closed*

lemma *subgraph-isomorphic-trans*[*trans*]:
assumes $G_1 \sqsubseteq G_2$ **and** $G_2 \sqsubseteq G_3$
shows $G_1 \sqsubseteq G_3$
⟨*proof*⟩

lemma *subgraph-isomorphic-post-iso-closed*: $[\![H \sqsubseteq G; G \simeq G']\!] \Longrightarrow H \sqsubseteq G'$
⟨*proof*⟩

lemmas *subgraph-isomorphic-post-closed* = *subgraph-isomorphic-post-iso-closed*

lemmas *subgraph-isomorphic-closed* = *subgraph-isomorphic-pre-closed* *subgraph-isomorphic-post-closed*

4.5 Density

The density of a graph is the quotient of the number of edges and the number of vertices of a graph.

definition *density* :: *ugraph* ⇒ *real* **where**
density G = *card* (*uedges* G) / *card* (*uverts* G)

The maximum density of a graph is the density of its densest nonempty subgraph.

definition *max-density* :: *ugraph* ⇒ *real* **where**
max-density G = *Lattices-Big.Max* (*density* ‘ *nonempty-subgraphs* G)

We prove some obvious results about the maximum density, such as that there is a subgraph which has the maximum density and that the (maximum) density is preserved by isomorphisms. The proofs are a bit complicated by the fact that most facts about *Max* require non-emptiness of the target set, but we need that anyway to get a value out of it.

lemma *subgraph-has-max-density*:
assumes *finite-graph* G **and** *nonempty-graph* G **and** *uwellformed* G
shows $\exists G'$. *density* G' = *max-density* G ∧ *subgraph* G' G ∧ *nonempty-graph* G' ∧ *finite-graph* G' ∧ *uwellformed* G'

⟨*proof*⟩

lemma *max-density-is-max*:
assumes *finite-graph* G **and** *finite-graph* G' **and** *nonempty-graph* G' **and** *uwellformed* G' **and** *subgraph* G' G
shows *density* $G' \leq$ *max-density* G
⟨*proof*⟩

lemma *max-density-gr-zero*:
assumes *finite-graph* G **and** *nonempty-graph* G **and** *uwellformed* G
shows $0 <$ *max-density* G
⟨*proof*⟩

lemma *isomorphic-density*:
assumes $G_1 \simeq G_2$
shows *density* G_1 = *density* G_2
⟨*proof*⟩

lemma *isomorphic-max-density*:
assumes $G_1 \simeq G_2$ **and** *nonempty-graph* G_1 **and** *nonempty-graph* G_2 **and** *finite-graph* G_1 **and** *finite-graph* G_2
shows *max-density* G_1 = *max-density* G_2
⟨*proof*⟩

4.6 Fixed selectors

In the proof of the main theorem in the lecture notes, the concept of a "fixed copy" of a graph is fundamental.

Let H be a fixed graph. A 'fixed selector' is basically a function mapping a set with the same size as the vertex set of H to a new graph which is isomorphic to H and its vertex set is the same as the input set.[1]

definition *is-fixed-selector* $H\,f = (\forall\, V.\ \mathit{finite}\ V \land \mathit{card}\ (\mathit{uverts}\ H) = \mathit{card}\ V \longrightarrow H \simeq f\ V \land \mathit{uverts}\ (f\ V) = V)$

Obviously, there may be many possible fixed selectors for a given graph. First, we show that there is always at least one. This is sufficient, because we can always obtain that one and use its properties without knowing exactly which one we chose.

lemma *ex-fixed-selector*:
assumes *uwellformed* H **and** *finite-graph* H
obtains f **where** *is-fixed-selector* $H\,f$
⟨*proof*⟩

lemma *fixed-selector-induced-subgraph*:
assumes *is-fixed-selector* $H\,f$ **and** *card* (*uverts* H) = *card* V **and** *finite* V

[1] We call such a selector *fixed* because its result is deterministic.

assumes *sub*: *subgraph* (*f V*) (*induced-subgraph V G*) **and** *V*: $V \subseteq$ *uverts G* **and** *G*: *uwellformed G*
shows $H \sqsubseteq G$
⟨*proof*⟩

end

5 Classes and properties of graphs

theory *Ugraph-Properties*
imports
Ugraph-Lemmas
Girth-Chromatic.Girth-Chromatic
begin

A "graph property" is a set of graphs which is closed under isomorphism.

type-synonym *ugraph-class* = *ugraph set*

definition *ugraph-property* :: *ugraph-class* ⇒ *bool* **where**
ugraph-property $C \equiv \forall\, G \in C.\ \forall\, G'.\ G \simeq G' \longrightarrow G' \in C$

abbreviation *prob-in-class* :: (*nat* ⇒ *real*) ⇒ *ugraph-class* ⇒ *nat* ⇒ *real* **where**
prob-in-class p c n ≡ *probGn p n* (λ*es*. *edge-space.edge-ugraph n es* ∈ *c*)

From now on, we consider random graphs not with fixed edge probabilities but rather with a probability function depending on the number of vertices. Such a function is called a "threshold" for a graph property iff

- for asymptotically *larger* probability functions, the probability that a random graph is an element of that class tends to 1 ("1-statement"), and
- for asymptotically *smaller* probability functions, the probability that a random graph is an element of that class tends to 0 ("0-statement").

definition *is-threshold* :: *ugraph-class* ⇒ (*nat* ⇒ *real*) ⇒ *bool* **where**
is-threshold c t ≡ *ugraph-property c* ∧ (∀ *p*. *nonzero-prob-fun p* ⟶
($p \ll t$ ⟶ *prob-in-class p c* ⟶ *0*) ∧
($t \ll p$ ⟶ *prob-in-class p c* ⟶ *1*))

lemma *is-thresholdI*[*intro*]:
assumes *ugraph-property c*
assumes ⋀*p*. ⟦ *nonzero-prob-fun p*; $p \ll t$ ⟧ ⟹ *prob-in-class p c* ⟶ *0*
assumes ⋀*p*. ⟦ *nonzero-prob-fun p*; $t \ll p$ ⟧ ⟹ *prob-in-class p c* ⟶ *1*
shows *is-threshold c t*
⟨*proof*⟩

end

6 The subgraph threshold theorem

theory *Subgraph-Threshold*
imports
  Ugraph-Properties
begin

lemma (**in** *edge-space*) *measurable-pred*[*measurable*]: *Measurable.pred P Q*
  ⟨*proof*⟩

This section contains the main theorem. For a fixed nonempty graph H, we consider the graph property of 'containing an isomorphic subgraph of H'. This is obviously a valid property, since it is closed under isomorphism.

The corresponding threshold function is

$$t(n) = n^{-\frac{1}{\rho'(H)}},$$

where ρ' denotes *max-density*.

definition *subgraph-threshold* :: *ugraph* ⇒ *nat* ⇒ *real* **where**
subgraph-threshold H n = *n powr* (−(*1* / *max-density H*))

theorem
  assumes *nonempty*: *nonempty-graph H* **and** *finite*: *finite-graph H* **and** *well-formed*: *uwellformed H*
  shows *is-threshold* {*G*. *H* ⊑ *G*} (*subgraph-threshold H*)
⟨*proof*⟩

end

References

[1] R. Diestel. *Graph Theory, 4th Edition*, volume 173 of *Graduate texts in mathematics*. Springer, 2012.

[2] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 5:17–61, 1960.

[3] S. Janson, T. Luczak, and A. Rucinski. *Random graphs*. John Wiley & Sons, 2011.

[4] L. Noschinski. A probabilistic proof of the girth-chromatic number theorem. *Archive of Formal Proofs*, Feb. 2012. http://isa-afp.org/entries/Girth_Chromatic.shtml, Formal proof development.