Stockholms universitet

Research Report

Department of Statistics

No. 2012:3

Some Multigraph Algorithms

Termeh Shafie

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Some Multigraph Algorithms

Termeh Shafie

Abstract

Several algorithms for generating and analyzing multigraphs under two different multigraph models are presented including the following: to find distributions of complexity measures in different random multigraphs, to analyze the local and global structure of multigraphs under different multigraph models using information theoretic tools based on entropy, and to test simple or composite hypotheses concerning random multigraphs.

Keywords: multigraph, algorithm, graph enumeration, complexity, multiplicity, entropy, information divergence, goodness-of-fit.

Introduction

There are many graph theoretic algorithms available in the literature. In this article, multigraph algorithms used in the articles by Frank and Shafie (2012), Shafie (2012a) and Shafie (2012b) are presented. Algorithms are given for generating multigraphs under two different multigraph models. The first model is random stub matching (RSM) where the edges are formed by randomly coupling pairs of stubs according to a fixed stub multiplicity or degree sequence. Thus, edge assignments to vertex pair sites are dependent. The second multigraph model is obtained by independent edge assignments (IEA) according to a common probability distribution over the sites. Two different methods for obtaining an approximate IEA model from an RSM model are also considered. The first method is obtained by assuming that the stubs are randomly generated and independently assigned to vertices, called independent stub assignments (ISA) and the second method of obtaining an approximate IEA model is to ignore the dependency between edges in the RSM model and assume independent edge assignments of stubs (IEAS).

Algorithms are also given for analyzing and testing different multigraph models using information theoretic tools based on entropy. In particular, algorithms are given for using the local and global distributions under RSM and IEA to calculate moments and entropies,

Department of Statistics, Stockholm University, S-106 91 Stockholm, termeh.shafie@stat.su.se

and for comparisons between distributions by information divergence. Further, special algorithms are developed for analyzing the complexity of multigraphs which is defined and quantified by the distribution of edge multiplicities.

All algorithms are developed by me and presented in a brief algorithmic style. They have been used for research work presented in the references, and details about concepts and notations used in the algorithms can be found there. Note that there might be more efficient alternatives available in the computer science literature. Such efficiency might be required in order to apply the methods to large multigraphs or to extensive calculations with many multigraphs. Since this has not been needed during the methodological development, no attempts have been made to get optimal algorithms.

List of Algorithms

Algorithm 1: Generating non-decreasing edge sequences for multigraphs with fixed number of vertices and edges

Input: Number of vertices n and number of edges m
Output: A list $\mathbf{Z}$ with all possible non-decreasing edge sequences

```
r ← binom(n+1, 2)
t ← 1
for i ← 1 to m do
    e(t, i) ← 1
k ← m
while k > 0 do
    while e(t, k) < r do
        t ← t + 1
        for i ← 1 to m do
            if i < k then
                e(t, i) ← e(t − 1, i)
            else if i ≥ k and e(t, i) + 1 ≤ r then
                e(t, i) ← e(t − 1, k) + 1
        k ← m
    if e(t, k) = r then
        k ← k − 1
Z ← e
for i ← 1 to n do
    for j ← 1 to n do
        if i ≤ j then
            A(i, j) = 1
        else
            A(i, j) = 0
foreach row in Z do
    foreach column c ← 1 to m do
        for i ← 1 to n do
            for j ← 1 to n do
                if A(i, j) > 0 then
                    a ← row index i
                    b ← column index j
                    recode c ← (a, b)
return Z
```

Algorithm 2: Complexity of multigraphs with fixed number of vertices and edges

Input: Number of vertices n and number of edges m, edge loops *allowed* or *forbidden*
Output: Properties of multigraphs including complexity sequences $\mathbf{r} = (r_0, \ldots, r_m)$ and summary complexity measure t for each possible non-decreasing edge sequence

1 Call **Algorithm 1** for e and $\mathbf{Z}$
2 $r \leftarrow \binom{n+1}{2}$
3 **foreach** *row* i *in* $\mathbf{Z}$ **do**
4   **for** $j \leftarrow 1$ ***to*** r **do**
5   Multiplicity sequence $\mathbf{m}(i, j) \leftarrow$ frequency of $e(i) = j$
6   **for** $j \leftarrow 0$ ***to*** m **do**
7   Complexity sequence $\mathbf{r}(i, j+1) \leftarrow$ frequency of $\mathbf{m}(i) = j$
8     $M(i) \leftarrow$ upper triangular matrix containing elements of $\mathbf{m}(i)$
9   Number of edge loops $m_1(i) \leftarrow$ sum over all diagonal elements $M(i)$
10   **if** $m_1(i) > 0$ **then**
11         $I_{m_1} \leftarrow 1$
12 **if** *forbidden* **then**
13   Remove all rows in $\mathbf{Z}$ where $I_{m_1} = 1$
14   Repeat steps 3-7
15 **foreach** *row* i *in* $\mathbf{Z}$ **do**
16     $mFac(i) \leftarrow$ product over the factorial of each element in $\mathbf{m}(i)$
17   Complexity summary measure $t(i) \leftarrow m_1(i) + \log_2 mFac(i)$
18   **for** $u \leftarrow 1$ ***to*** n **do**
19   Degree sequence $\mathbf{d}(i) \leftarrow$ frequency of $\mathbf{Z}(i) = u$
20 **return** List with columns containing $\mathbf{Z}$, $\mathbf{d}$, $\mathbf{m}$, m_1, $\mathbf{r}$, t

Algorithm 3: Generating non-decreasing permutations of a stub multiplicity sequence

Input: Degree or stub multiplicity sequence $\mathbf{d} = (d_1, d_2, \ldots, d_n)$
Output: A list $\mathbf{S}$ with all non-decreasing permutations of a stub sequence

Number of vertices $n \leftarrow$ number of columns in $\mathbf{d}$
Number of edges $m \leftarrow$ half of the sum of all element values in $\mathbf{d}$
$t \leftarrow 1$
$\mathbf{s}(t) \leftarrow [1^{d_1}\ 2^{d_2}\ \cdots\ n^{d_n}]$
$k \leftarrow m - 1$
while $k > 0$ **do**
    $W_k(t) \leftarrow [1^{d_1(t,k)} \cdots n^{d_n(t,k)}]$, an ordered sequence of vertices in the edges $(e_k(t), \ldots, e_m(t))$
    Try to re-order $W_k(t)$ as a non-decreasing sequence of edges $(f_k(t), \ldots, f_m(t))$ above $e_k(t)$ so that $f_k(t) \leftarrow e_k(t) + 1 \leq f_{k+1}(t) \leq \cdots \leq f_m(t)$, where $e_k(t) + 1$ means that its second vertex is increased by 1
    if *re-order possible* **then**
        $\mathbf{s}(t+1) \leftarrow [e_1(t), \ldots, e_{k-1}(t), f_k(t), \ldots, f_m(t)]$
        $t \leftarrow t + 1$
    else
        $k \leftarrow k - 1$
$\mathbf{S} \leftarrow \mathbf{s}$
return $\mathbf{S}$

Algorithm 4: Complexity distribution and multigraph distribution for multigraphs under RSM

Input: Degree or stub multiplicity sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$
Output: Lists including complexity distributions P_t and multigraph distributions $P_{\mathbf{Z}}$

Number of vertices $n \leftarrow$ number of columns in $\mathbf{d}$
Number of edges $m \leftarrow$ half of the sum of all element values in $\mathbf{d}$
Call **Algorithm 3** for $\mathbf{S} \leftarrow$ list of all permutations of the stub sequence
$\mathbf{Z} \leftarrow \mathbf{S}$
Follow steps 2-19 in **Algorithm 2** to find $\mathbf{m}$, m_1, $\mathbf{r}$, t
foreach *row i in* $\mathbf{Z}$ **do**
    if *any element in* $\mathbf{m}(i) \geq 2$ **then**
        Multiple edge indicator $I_{m_2}(i) \leftarrow 1$
        $x \leftarrow$ a vector containing elements in $\mathbf{m}(i) \geq 2$
        $C \leftarrow$ product over the factorial of each element in x
    else
        Multiple edge indicator $I_{m_2}(i) \leftarrow 0$
    if $I_{m_2}(i) = 1$ **then**
        Total number of possible shift permutations between edge pairs $SPB(i) \leftarrow m!/C$
    else
        Total number of possible shift permutations between edge pairs $SPB(i) \leftarrow m!$
    if $m_1(i) = m$ **then**
        Total number of possible shift permutations within edge pairs $SPW(i) \leftarrow 1$
    else
        Total number of possible shift permutations within edge pairs $SPW(i) \leftarrow 2^{m-m_1(i)}$
    Total number of permutations of each edge sequence $K_{\mathbf{Z}}(i) \leftarrow SPB(i) \cdot SPW(i)$
Total number of multigraphs $K_{\mathbf{d}} \leftarrow$ sum over all elements in $K_{\mathbf{Z}}$
foreach *row i in* $\mathbf{Z}$ **do**
    Probability of multigraph $P_{\mathbf{Z}}(i) \leftarrow K_{\mathbf{Z}}(i)/K_{\mathbf{d}}$
$t_{UNI} \leftarrow$ unique values of t
foreach *row i in t* **do**
    Number of complexity value $K_t(i) \leftarrow$ frequency of $t_{UNI} = t(i)$
    Probability of complexity value $P_t(i) \leftarrow$ sum over all $P_{\mathbf{Z}}$ where $t_{UNI} = t(i)$
return List with columns containing $\mathbf{Z}$, $P_{\mathbf{Z}}$, $K_{\mathbf{Z}}$, $\mathbf{m}$, m_1, $\mathbf{r}$ and list with columns containing t_{UNI}, K_t, P_t

Algorithm 5: Entropies of trivariate edge multiplicity distributions under RSM and IEA

Input: Number of edges m, degrees d_i and d_j at vertices i and j

Output: The entropies h, entropy upper bounds $Maxh$, and entropy approximations $Apxh$ of $P((m_{ii}, m_{jj}, m_{ij})) = (u, v, w))$ under RSM and IEA

$a \leftarrow d_i,\ \ b \leftarrow d_j,\ \text{ and }\ c \leftarrow \min(a, b)$
$i \leftarrow 0$
for $w \leftarrow 0$ ***to*** c **do**
  for $u \leftarrow 0$ ***to*** $\lfloor (a-w)/2 \rfloor$ **do**
    for $v \leftarrow 0$ ***to*** $\lfloor (b-w)/2 \rfloor$ **do**
      if $(m - a - b + u + v + w) \geq 0$ **then**
        $i = i + 1$
        $UVW(i) \leftarrow [u\ v\ w]$
        $P(i) \leftarrow \frac{m!\ a!\ b!\ 2^{(a+b-2u-2v-w)}(2m-a-b)!}{u!\ v!\ w!(a-2u-w)!\ (b-2v-w)!\ (m-a-b+u+v+w)!\ (2m)!}$

foreach *row* i *in* P **do**
  if $P(i) > 0$ **then**
    $\phi(i) \leftarrow -P(i) \log_2 P(i)$
  else
    $\phi(i) \leftarrow 0$

Entropy $h_{RSM} \leftarrow$ sum over all elements in ϕ
Max entropy $Maxh_{RSM} \leftarrow \log_2$ of number of rows in UVW
$Cov \leftarrow$ covariance matrix of UVW
Entropy approximation $Apxh_{RSM} \leftarrow \log_2 \left(\sqrt{2\pi e Cov}\right)$
$Q_{aa} \leftarrow \frac{a(a-1)}{2m(2m-1)},\ \ Q_{bb} \leftarrow \frac{b(b-1)}{2m(2m-1)},\ \text{ and }\ Q_{ab} \leftarrow \frac{2ab}{2m(2m-1)}$
$Q_c \leftarrow (1 - Q_{aa} - Q_{bb} - Q_{ab})$
$multprobs \leftarrow [Q_{aa}\ Q_{bb}\ Q_{ab}\ Q_c]$
$U \leftarrow 0$ to m, $V \leftarrow 0$ to m, and $W \leftarrow 0$ to m
Produce three-dimensional coordinate arrays where the output coordinate arrays u, v, and w contain copies of the grid vectors U, V, and W, respectively
$x \leftarrow m - (u + v + w)$
$\mathbf{X} \leftarrow [u\ v\ w\ x]$
Remove rows in $\mathbf{X}$ that have negative elements and check that there are $\binom{m+3}{3}$ rows left
for $i \leftarrow 0$ ***to*** $\binom{m+3}{3}$ **do**
  $B(i+1) \leftarrow$ the pdf for the multinomial distribution with probabilities *multprobs* evaluated at each row $\mathbf{X}(i+1)$

for $i \leftarrow 0$ ***to*** $\binom{m+3}{3}$ **do**
  if $B(i) > 0$ **then**
    $\phi(i) \leftarrow -B(i) \log_2 B(i)$
  else
    $\phi(i) \leftarrow 0$

Entropy $h_{IEA} \leftarrow$ sum over all elements in ϕ
Max entropy $Maxh_{IEA} \leftarrow \log_2 \binom{m+3}{3}$
Entropy approximation $Apxh_{IEA} \leftarrow \log_2 \left(\sqrt{(2\pi e m)^3 Q_{aa} Q_{bb} Q_{ab} Q_c}\right)$
return h_{RSM}, $Maxh_{RSM}$, $Apxh_{RSM}$, h_{IEA}, $Maxh_{IEA}$, $Apxh_{IEA}$

Algorithm 6: Entropies and moments of marginal loop distributions under RSM and IEA

Input: Number of edges m

Output: The entropies h and entropy approximations $Apxh$ of $P(m_{ii} = v)$ under RSM and IEA, together with expected values and variances, respectively

for $i \leftarrow 2$ ***to*** $2m$ **do**
$deg(i) = i$
foreach *row i in deg* **do**
$d = deg(i)$
for $v \leftarrow 0$ ***to*** $\lfloor d/2 \rfloor$ **do**
if $(m - d + v) < 0$ **then**
$P(i, v+1) \leftarrow 0$
else
$P(i, v+1) \leftarrow \big(m!\ 2^{d-2v}\ d!\ (2m-d)!\big) \,/\, (v!\ (d-2v)!\ (m-d+v)!\ 2m!)$
$Q \leftarrow d(d-1)/(2m(2m-1))$
for $v \leftarrow 0$ ***to*** m **do**
$B(i, v+1) \leftarrow$ the pdf for the binomial distribution with parameters Q and m evaluated at point v
$\mu(i) \leftarrow d(d-1)/(2(2m-1))$
$\sigma^2(i) \leftarrow \mu(i)(1 - \mu(i)/m)$
$\Delta(i) \leftarrow (d(d-1)(d-2)(d-3)) \,/\, (4(2m-1)(2m-3)) - \big(\mu(i)^2((m-1)/m)\big)$
$Var \leftarrow \sigma^2(i) + \Delta(i)$
foreach *row i in P* **do**
foreach *column j in P* **do**
if $P(i,j) > 0$ **then**
$\phi(i,j) \leftarrow -P(i,j)\log_2 P(i,j)$
else
$\phi(i,j) \leftarrow 0$
Entropy $h_{RSM}(i) \leftarrow$ sum over all columns in $\phi(i)$
foreach *row i in B* **do**
foreach *column j in B* **do**
if $B(i,j) > 0$ **then**
$\phi(i,j) \leftarrow -B(i,j)\log_2 B(i,j)$
else
$\phi(i,j) \leftarrow 0$
Entropy $h_{IEA}(i) \leftarrow$ sum over all columns in $\phi(i)$
foreach *row i in deg* **do**
Entropy approximation RSM $Apxh_{RSM}(i) \leftarrow \log_2\left(\sqrt{2\pi e Var(i)}\right)$
Entropy approximation IEA $Apxh_{IEA}(i) \leftarrow \log_2\left(\sqrt{2\pi e \sigma^2(i)}\right)$
return List with columns containing deg, h_{RSM}, h_{IEA}, $Apxh_{RSM}$, $Apxh_{IEA}$, μ, σ^2, Δ, Var

Algorithm 7: Entropies and moments of marginal non-loop distributions under RSM and IEA

Input: Number of edges m

Output: The entropies h and entropy approximations $Apxh$ of $P(m_{ij} = w)$ under RSM and IEA, together with expected values and variances, respectively

```
for i ← 2 to m do
    for j ← i to 2m − i do
        Each row in deg ← [i j]
foreach row i in deg do
    d_i ← deg(i, 1)
    d_j ← deg(i, 2)
    Call Algorithm 5 for P
    a ← d_i,  b ← d_j  and  c ← min(a, b)
    for w ← 0 to c do
        P_w(i, w + 1) ← sum over a to ⌊(a − w)/2⌋ and b to ⌊(b − w)/2⌋ in P
    Q ← 2ab/(2m(2m − 1))
    for w ← 0 to m do
        B_w(i, w + 1) ← the pdf for the binomial distribution with parameters Q and m evaluated at
        point w
    μ(i) ← 2ab/(2(2m − 1))
    σ^2(i) ← μ(i)(1 − μ(i)/m)
    Δ(i) ← (ab(a − 1)(b − 1)) / ((2m − 1)(2m − 3)) − (μ(i)^2((m − 1)/m))
    Var ← σ^2(i) + Δ(i)
foreach row i in P_w do
    foreach column j in P_w do
        if P_w(i, j) > 0 then
            φ(i, j) ← −P_w(i, j) log_2 P_w(i, j)
        else
            φ(i, j) ← 0
    Entropy h_RSM(i) ← sum over all columns in φ(i)
foreach row i in B_w do
    foreach column j in B_w do
        if B_w(i, j) > 0 then
            φ(i, j) ← −B_w(i, j) log_2 B_w(i, j)
        else
            φ(i, j) ← 0
    Entropy h_IEA(i) ← sum over all columns in φ(i)
foreach row i in D do
    Entropy approximation RSM Apxh_RSM(i) ← log_2(sqrt(2πeVar(i)))
    Entropy approximation IEA Apxh_IEA(i) ← log_2(sqrt(2πeσ^2(i)))
return List with columns containing deg, h_RSM, h_IEA, Apxh_RSM, Apxh_IEA, μ, σ^2, Δ, Var
```

Algorithm 8: Information divergence between trivariate edge multiplicity distributions under RSM and IEA

Input: Number of edges m, degrees d_i and d_j at vertices i and j
Output: The information divergence D between $P((m_{ii}, m_{jj}, m_{ij})) = (u, v, w))$ under RSM and IEA

Call **Algorithm 5** for UVW, P and *multprobs*
foreach *row* i *in* UVW **do**
    $UVW(i, 4) \leftarrow (m - UVW(i, 1) - UVW(i, 2) - UVW(i, 3))$
foreach *row* i *in* UVW **do**
    $B(i) \leftarrow$ the pdf for the multinomial distribution with probabilities *multprobs* evaluated at each row $UVW(i)$
foreach *row* i *in* UVW **do**
    Weighted log-likelihood ratio $LLR(i) \leftarrow P_{uvw}(i) \log_2 (P(i)/B(i))$
Divergence $D \leftarrow$ sum over all elements in LLR
return D

Algorithm 9: Information divergence between marginal loop multiplicity distributions under RSM and IEA

Input: Number of edges m
Output: The information divergence D between $P(m_{ii} = v)$ under RSM and IEA

Call **Algorithm 6** for *deg*, P and B
foreach *row* i *in* P **do**
    foreach *column* j *in* P **do**
        if $P(i, j) > 0$ **then**
            Weighted log-likelihood ratio $LLR(j) \leftarrow P(i, j) \log_2 (P(i, j)/B(i, j))$
        else
            $LLR(j) \leftarrow 0$
    Divergence $D(i) \leftarrow$ sum over all elements $LLR(j)$
return List with columns containing *deg*, D

Algorithm 10: Information divergence between marginal non-loop multiplicity distributions under RSM and IEA

Input: Number of edges m
Output: The information divergence D between $P(m_{ij} = w)$ under RSM and IEA

Call **Algorithm 7** for *deg*, P_w and B_w
foreach *row* i *in* P_w **do**
    foreach *column* j *in* P_w **do**
        if $P_w(i, j) > 0$ **then**
            Weighted log-likelihood ratio $LLR(j) \leftarrow P_w(i, j) \log_2 (P_w(i, j)/B_w(i, j))$
        else
            $LLR(j) \leftarrow 0$
    Divergence $D(i) \leftarrow$ sum over all elements $LLR(j)$
return List with columns containing *deg*, D

Algorithm 11: Entropies of and information divergence between distributions of multigraphs under RSM and IEA

Input: Degree or stub multiplicity sequence $\mathbf{d} = (d_1, d_2, \ldots, d_n)$

Output: The entropies h, entropy upper bounds $Maxh$, entropy approximation $Apxh$ and information divergence D of and between the multigraph distributions under RSM and IEA

Number of vertices $n \leftarrow$ number of columns in $\mathbf{d}$

Number of edges $m \leftarrow$ half of the sum of all element values in $\mathbf{d}$

Call **Algorithm 4** for $\mathbf{m}$, $P_{\mathbf{Z}}$ and $K_{\mathbf{d}}$

foreach *row* i *in* $P_{\mathbf{Z}}$ **do**

  if $P_{\mathbf{Z}}(i) > 0$ **then**

    $\phi(i) \leftarrow -P_{\mathbf{Z}}(i) \log_2 P_{\mathbf{Z}}(i)$

  else

    $\phi(i) \leftarrow 0$

Entropy $h_{RSM} \leftarrow$ sum over all elements in ϕ

Max entropy $Maxh_{RSM} \leftarrow \log_2 K_{\mathbf{d}}$

$CovRSM \leftarrow$ covariance matrix of $\mathbf{m}$

Entropy approximation $Apxh_{RSM} \leftarrow \log_2 \left(\sqrt{det(2\pi e CovRSM)} \right)$

for $i \leftarrow 0$ ***to*** n **do**

  for $j \leftarrow 0$ ***to*** n **do**

    if $i = j$ **then**

      $\mathbf{Q}(i,j) \leftarrow (\mathbf{d}(i)(\mathbf{d}(i) - 1))/(2m(2m-1))$

    else if $i < j$ **then**

      $\mathbf{Q}(i,j) \leftarrow 2\mathbf{d}(i)\mathbf{d}(j)/(2m(2m-1))$

    else

      $\mathbf{Q}(i,j) \leftarrow 0$

$Q \leftarrow$ vector containing the upper triangular elements $i \leq j$ of $\mathbf{Q}(i,j)$

$r \leftarrow \binom{n+1}{2}$

for $i \leftarrow 1$ ***to*** $r-1$ **do**

  for $j \leftarrow 1$ ***to*** $r-1$ **do**

    if $i = j$ **then**

      $CovIEA(i,j) \leftarrow mQ(i)(1 - Q(i))$

    else

      $CovIEA(i,j) \leftarrow -mQ(i)Q(j)$

Max entropy $Maxh_{IEA} \leftarrow \log_2 \binom{m+r-1}{m}$

Entropy approximation $Apxh_{IEA} \leftarrow \log_2 \left(\sqrt{det(2\pi e CovIEA)} \right)$

for $i \leftarrow 1$ ***to*** $\mathbf{m}$ **do**

  $P_{IEA}(i) \leftarrow$ the pdf for the multinomial distribution with probabilities Q evaluated at each row $\mathbf{m}(i)$

foreach *row* i *in* $P_{\mathbf{Z}}$ **do**

  Weighted log-likelihood ratio $LLR(i) \leftarrow P_{\mathbf{Z}}(i) \log_2 (P_{\mathbf{Z}}(i)/P_{IEA}(i))$

Divergence $D \leftarrow$ sum over all elements in LLR

return h_{RSM}, $Maxh_{RSM}$, $Apxh_{RSM}$, $Maxh_{IEA}$, $Apxh_{IEA}$, D

Algorithm 12: Entropy approximations of distributions of multigraphs under RSM

Input: Degree or stub multiplicity sequence $\mathbf{d} = (d_1, d_2, \ldots, d_n)$
Output: Entropy H under RSM, and the two entropy approximations H^* and H^{**} based on expected entropy and asymptotic entropy under ISA

Number of vertices $n \leftarrow$ number of columns in $\mathbf{d}$
Number of edges $m \leftarrow$ half of the sum of all element values in $\mathbf{d}$
Call **Algorithm 11** for h_{RSM}
$H \leftarrow h_{RSM}$
$prod \leftarrow$ product over all the elements in $\mathbf{d}$
if $n > 2$ **then**
    $H^* \leftarrow \log_2 \left(\sqrt{(2\pi e m)^{\binom{n}{2}} 2^{\binom{n-1}{2}} prod^n (1/2m)^{n^2}} \right)$
else
    $H^* \leftarrow \log_2 \left(\sqrt{(2\pi e m)^{\binom{n}{2}} prod^n (1/2m)^{n^2}} \right)$
$\mathbf{p} \leftarrow$ vector containing each element in $\mathbf{d}$ divided by $2m$
for $i \leftarrow 1$ ***to*** n **do**
    $x(i) \leftarrow -\mathbf{p}(i) \log_2 \mathbf{p}(i)$
$h_{\mathbf{p}} \leftarrow$ sum over all elements in x
$n_c \leftarrow 2^{h_{\mathbf{p}}}$
$r_c \leftarrow n_c^2 2^{-(n_c-1)/n_c}$
$a^{**} \leftarrow \log_2 \left(\sqrt{(2\pi e)^{r_c-1} (n_c^{n_c}) / (4\pi e)^{n_c-1} r_c^{r_c}} \right)$
$b^{**} \leftarrow (r_c - n_c)/2$
$H^{**} \leftarrow a^{**} + b^{**} \log_2(m)$
return H, H^*, H^{**}

Algorithm 13: Approximations of the probability that an RSM multigraph is simple

Input: Degree or stub multiplicity sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$
Output: The probability that an RSM multigraph is simple P_{RSM}, together with the three approximation P_{1Asymp}, P_{2Asymp} and $P_{Poisson}$

```
Number of vertices n ← number of columns in d
Number of edges m ← half of the sum of all element values in d
Call Algorithm 4 for m, m_1, I_{m_2} and K_d
foreach row i in m do
    if m_1(i) = 0 and I_{m_2}(i) = 0 then
        simple(i) ← 1
    else
        simple(i) ← 0
P_RSM ← sum over all elements in simple divided by K_d
S ← sum over all squared elements in d
P_1Asymp ← exp(−1/4(S/2m)^2 + 1/4)
for i ← 0 to n do
    for j ← 0 to n do
        if i = j then
            L(i, j) ← d(i)(d(i) − 1)/2m
        else if i < j then
            L(i, j) ← sqrt(d(i)(d(i) − 1)d(j)(d(j) − 1)/2m)
        else
            L(i, j) ← 0
X ← L − log(1 + L)
Y ← matrix with elements above the diagonal in X
sum1 ← sum over the diagonal in L
sum2 ← sum over all rows and columns in Y
P_2Asymp ← exp(−1/2(sum1) − sum2)
for i ← 0 to n do
    for j ← 0 to n do
        if i = j then
            Q(i, j) ← (d(i)(d(i) − 1))/(2m(2m − 1))
        else if i < j then
            Q(i, j) ← 2d(i)d(j)/(2m(2m − 1))
        else
            Q(i, j) ← 0
for i ← 0 to n do
    for j ← 0 to n do
        if i < j then
            bin(i, j) ← the pdf for the binomial distribution with parameters Q(i, j) and m
            evaluated at point 1
λ ← sum over all columns and rows in bin
P_Poisson ← the pdf of the Poisson distribution with parameter λ evaluated at point m
return P_RSM, P_1Asymp, P_2Asymp, P_Poisson
```

Algorithm 14: Distribution of the multiplicity sequence under an RSM, IEAS or ISA model

Input: Model $RSM(\mathbf{d})$, $IEAS(\mathbf{d})$ or $ISA(\mathbf{p})$, where $\mathbf{d}$ and $\mathbf{p}$ are specified
Output: Multiplicity sequences $\mathbf{m}$ and their probabilities *probs* under the specified model

```
r ← binom(n+1, 2)
if Model=RSM(d) then
    Number of vertices n ← number of columns in d
    Number of edges m ← half of the sum of all element values in d
    Call Algorithm 4 for m and P_Z
    probs ← P_Z
else if Model=IEAS(d) then
    Number of vertices n ← number of columns in d
    Number of edges m ← half of the sum of all element values in d
    Call Algorithm 11 for Q
    Q ← vector containing the upper triangular elements i ≤ j of Q(i, j)
    m ← binom(m+r−1, m) rows of possible multiplicity sequences under IEAS
    foreach row i in m do
        probs(i) ← the pdf for the multinomial distribution with probabilities Q evaluated at each
        row m(i)
else if Model=ISA(p) then
    Number of vertices n ← number of columns in p
    Number of edges m ← half of the sum of all element values in p
    for i ← 0 to n do
        for j ← 0 to n do
            if i = j then
                Q(i, j) ← p(i)^2
            else if i < j then
                Q(i, j) ← 2p(i)p(j)
            else
                Q(i, j) ← 0
    Q ← vector containing the upper triangular elements i ≤ j of Q(i, j)
    m ← binom(m+r−1, m) rows of possible multiplicity sequences under ISA
    foreach row i in m do
        probs(i) ← the pdf for the multinomial distribution with probabilities Q evaluated at each
        row m(i)
return m, probs
```

Algorithm 15: Statistical tests of a simple IEA hypothesis

Input: Model $RSM(\mathbf{d})$, $IEAS(\mathbf{d})$ or $ISA(\mathbf{p})$, and hypothesis $IEAS(\mathbf{d_0})$ or $ISA(\mathbf{p_0})$ where $\mathbf{d}$, $\mathbf{d_0}$, $\mathbf{p}$ and $\mathbf{p_0}$ are specified

Output: Outcomes of the Pearson goodness-of-fit statistic S with probabilities P_S, and outcomes of the divergence statistic T with probabilities P_T

1 Call **Algorithm 14** for $\mathbf{m}$ and *probs* according to input model
2 **if** *Hypothesis*=$IEAS(\mathbf{d_0})$ **then**
3     $\mathbf{d} \leftarrow \mathbf{d_0}$
4   Call **Algorithm 11** for $\mathbf{Q}$
5     $Q \leftarrow$ vector containing the upper triangular elements $i \leq j$ of $\mathbf{Q}(i,j)$
6     $\mathbf{m} \leftarrow \binom{m+r-1}{m}$ rows of possible multiplicity sequences under IEAS
7   **foreach** *row* i *in* $\mathbf{m}$ **do**
8         $O_S(i) \leftarrow \mathbf{m}(i)$, $E_S(i) \leftarrow mQ$
9   **foreach** *column* j *in* $\mathbf{m}$ **do**
10   **if** $E_S(j) = 0$ **then**
11                 $x(j) \leftarrow 0$
12   **else**
13                 $x(j) \leftarrow (O_S(j) - E_S(j))^2 / E_S(j)$
14         $S(i) \leftarrow$ sum over all columns in x
15     $S_{UNI} \leftarrow$ unique values of S
16   **foreach** *row* i *in* S **do**
17         $P_S(i) \leftarrow$ sum over all *probs* where $S_{UNI} = S(i)$
18   **foreach** *row* i *in* $\mathbf{m}$ **do**
19         $O_D(i) \leftarrow \mathbf{m}(i)$, $E_D(i) \leftarrow mQ$
20   **foreach** *column* j *in* $\mathbf{m}(i)$ **do**
21   **if** $O_D(j) > 0$ ***and*** $E_D(j) > 0$ **then**
22                 $x(j) \leftarrow (O_D(j)/m) \log_2(O_D(j)/E_D(j))$
23   **else**
24                 $x(j) \leftarrow 0$
25         $D(i) \leftarrow$ sum over all columns in x
26         $T(i) \leftarrow 2mD(i)/\log_2(e)$
27     $T_{UNI} \leftarrow$ unique values of T
28   **foreach** *row* i *in* T **do**
29         $P_T(i) \leftarrow$ sum over all *probs* where $T_{UNI} = T(i)$
30 **else if** *Hypothesis*=$ISA(\mathbf{p_0})$ **then**
31   **for** $i \leftarrow 0$ ***to*** n **do**
32   **for** $j \leftarrow 0$ ***to*** n **do**
33   **if** $i = j$ **then**
34                 $\mathbf{Q}(i,j) \leftarrow \mathbf{p_0}(i)^2$
35   **else if** $i < j$ **then**
36                 $\mathbf{Q}(i,j) \leftarrow 2\mathbf{p_0}(i)\mathbf{p_0}(j)$
37   **else**
38                 $\mathbf{Q}(i,j) \leftarrow 0$
39   Repeat steps 5-29
40 **return** S_{UNI}, P_S, T_{UNI}, P_T

Algorithm 16: Statistical tests of a composite IEAS hypothesis

Input: Model $RSM(\mathbf{d})$ $IEAS(\mathbf{d})$ or $ISA(\mathbf{p})$ and hypothesis $IEAS$
Output: Outcomes of the Pearson goodness-of-fit statistic S with probabilities P_S, and outcomes of the divergence statistic T with probabilities P_T

Call **Algorithm 14** for $\mathbf{m}$ and *probs* according to input model
foreach *row* i *in* $\mathbf{m}$ **do**
  $M \leftarrow$ upper triangular matrix containing the elements in $\mathbf{m}(i)$
  $M \leftarrow M + M'$
  $\mathbf{d}_{EST}(i) \leftarrow$ sum over all rows (or columns) in M
  $\mathbf{d} \leftarrow \mathbf{d}_{EST}(i)$
  Call **Algorithm 11** for $\mathbf{Q}$
  $Q \leftarrow$ vector containing the upper triangular elements $i \leq j$ of $\mathbf{Q}(i, j)$
  $O_S(i) \leftarrow \mathbf{m}(i)$
  $E_S(i) \leftarrow mQ$
  foreach *column* j *in* $\mathbf{m}$ **do**
    if $E_S(j) = 0$ **then**
      $x(j) \leftarrow 0$
    else
      $x(j) \leftarrow (O_S(j) - E_S(j))^2 / E_S(j)$
  $S(i) \leftarrow$ sum over all columns in x
  $O_D(i) \leftarrow \mathbf{m}(i)$
  $E_D(i) \leftarrow mQ$
  foreach *column* j *in* $\mathbf{m}(i)$ **do**
    if $O_D(j) > 0$ ***and*** $E_D(j) > 0$ **then**
      $x(j) \leftarrow (O_D(j)/m) \log_2(O_D(j)/E_D(j))$
    else
      $x(j) \leftarrow 0$
  $D(i) \leftarrow$ sum over all columns in x
  $T(i) \leftarrow 2mD(i)/\log_2(e)$
  Clear M,$\mathbf{d}$, $\mathbf{Q}$, Q
$S_{UNI} \leftarrow$ unique values of S
foreach *row* i *in* S **do**
  $P_S(i) \leftarrow$ sum over all *probs* where $S_{UNI} = S(i)$
$T_{UNI} \leftarrow$ unique values of T
foreach *row* i *in* T **do**
  $P_T(i) \leftarrow$ sum over all *probs* where $T_{UNI} = T(i)$
return S_{UNI}, P_S, T_{UNI}, P_T

Algorithm 17: Statistical tests of a composite ISA hypothesis

Input: Model $RSM(\mathbf{d})$ $IEAS(\mathbf{d})$ or $ISA(\mathbf{p})$ and hypothesis ISA

Output: Outcomes of the Pearson goodness-of-fit statistic S with probabilities P_S, and outcomes of the divergence statistic T with probabilities P_T

Call **Algorithm 14** for $\mathbf{m}$ and *probs* according to input model

foreach *row* i *in* $\mathbf{m}$ **do**

  $M \leftarrow$ upper triangular matrix containing the elements in $\mathbf{m}(i)$

  $M \leftarrow M + M'$

  $\mathbf{d}_{EST}(i) \leftarrow$ sum over all rows (or columns) in M

  $\mathbf{p}_{EST}(i) \leftarrow$ each element in $\mathbf{d}_{EST}(i)$ divided by $2m$

  for $i \leftarrow 0$ ***to*** n **do**

    for $j \leftarrow 0$ ***to*** n **do**

      if $i = j$ **then**

        $\mathbf{Q}(i,j) \leftarrow \mathbf{p}_{EST}(i)^2$

      else if $i < j$ **then**

        $\mathbf{Q}(i,j) \leftarrow 2\mathbf{p}_{EST}(i)\mathbf{p}_{EST}(j)$

      else

        $\mathbf{Q}(i,j) \leftarrow 0$

  $Q(i) \leftarrow$ the upper triangular elements $i \leq j$ of $\mathbf{Q}(i,j)$

  Clear M, $\mathbf{Q}$

foreach *row* i *in* $\mathbf{m}$ **do**

  $Q \leftarrow Q(i)$

  Repeat steps 9-25 in**Algorithm 16**

Repeat steps 27-32 in **Algorithm 16**

return S_{UNI}, P_S, T_{UNI}, P_T

References

Frank, O. and Shafie, T. (2012), Complexity of Families of Multigraphs, to appear in *JSM Proceedings*, Section on Statistical Graphics, Alexandria, VA: American Statistical Association.

Shafie, T. (2012a), Random Stub Matching Models of Multigraphs, *Research Report 2012:1*, Department of Statistics, Stockholm University.

Shafie, T. (2012b), Statistical Analysis of Multigraphs, *Research Report 2012:2*, Department of Statistics, Stockholm University.