Hyperbolic graph generator☆

Rodrigo Aldecoa [a,*], Chiara Orsini [b], Dmitri Krioukov [a,c]

[a] *Northeastern University, Department of Physics, Boston, MA, USA*
[b] *Center for Applied Internet Data Analysis, University of California San Diego (CAIDA/UCSD), San Diego, CA, USA*
[c] *Northeastern University, Department of Mathematics, Department of Electrical & Computer Engineering, Boston, MA, USA*

ARTICLE INFO

ABSTRACT

Networks representing many complex systems in nature and society share some common structural properties like heterogeneous degree distributions and strong clustering. Recent research on network geometry has shown that those real networks can be adequately modeled as random geometric graphs in hyperbolic spaces. In this paper, we present a computer program to generate such graphs. Besides real-world-like networks, the program can generate random graphs from other well-known graph ensembles, such as the soft configuration model, random geometric graphs on a circle, or Erdős–Rényi random graphs. The simulations show a good match between the expected values of different network structural properties and the corresponding empirical values measured in generated graphs, confirming the accurate behavior of the program.

Program summary

Program title: Hyperbolic graph generator

Catalogue identifier: AEXC_v1_0

Program summary URL: http://cpc.cs.qub.ac.uk/summaries/AEXC_v1_0.html

Program obtainable from: CPC Program Library, Queen's University, Belfast, N. Ireland

Licensing provisions: GNU General Public License, version 3

No. of lines in distributed program, including test data, etc.: 101190

No. of bytes in distributed program, including test data, etc.: 771660

Distribution format: tar.gz

Programming language: C++.

Computer: Any.

Operating system: Any.

Classification: 6.3, 4.13, 23.

Nature of problem: Generation of graphs in hyperbolic spaces.

Solution method: Implementation based on analytical equations.

Additional comments: Can be used as a command-line tool or installed as a library to support more complex software.

Running time: Depends on the number of nodes. A few seconds for the graph in the example provided.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The interactions between components of a complex system are often represented as a network. This modeling allows for rigorous mathematical treatment, and broadens our understanding of the system [1]. Many real networks possess common structural patterns, including, in the first place, heterogeneous (often

Table 1
Regimes in the model.

γ \ T	0	$(0, \infty)$	∞
$[2, \infty)$	Hyperbolic RGGs	Soft hyperbolic RGGs	Soft configuration model
∞	Spherical RGGs	Soft spherical RGGs	Erdős–Rényi

power-law) distributions of node degrees [2], and strong clustering, i.e., higher numbers of triangular subgraphs than predicted by classical random graph models [3]. Recently introduced geometric graph models, based on the assumption that nodes in real networks are embedded in latent hyperbolic spaces [4,5], reproduce these common structural properties of real networks. Furthermore, these hyperbolic graphs replicate dynamical processes on top of real networks [6] and accurately predict missing links in them [7].

In this work we present a program to generate random hyperbolic graphs. This software implements and extends the network model introduced in [4]. Nodes are randomly sprinkled on a hyperbolic disk, and the probability of the existence of an edge (the connection probability) between two nodes is a function of the distance between the nodes in the hyperbolic space. Thus generated graphs have strong clustering, and node degree distributions in them are power laws. Moreover, other popular and well-studied random graph ensembles, namely the soft configuration model (SCM) [8], (soft) random geometric graphs (RGGs) on a circle [9,10], and Erdős–Rényi (ER) random graphs [11], appear as degenerate regimes in the model. Table 1 shows all the model regimes, the total of six. Each regime is defined by the values of only two parameters: γ, which is the expected exponent of the power-law degree distribution, and temperature T, the parameter controlling the strength of clustering in the network.

Researchers in different disciplines may benefit from the use of random hyperbolic graphs in their work. Yet the full implementation of the model and all its regimes is a tricky business, which involves dealing with some delicate details, due to a variety of internal parameters and their interactions over the six regimes. In Section 2–4 we describe the implementation details of the model, including how all the parameters are calculated in each regime. A good match between the values of expected graph properties and their observed values in generated graphs is confirmed in Section 5.

2. Input parameters and coordinates

2.1. Input parameters

The program input parameters are the number of nodes N, the target expected average degree $\bar{k}$ of the network, the target expected power-law exponent γ of the degree distribution, and temperature T. The combination of γ and T values will define the graph ensemble from which generated networks are sampled (Table 1).

Given the input parameters, the graph generation process consists of three steps:

1. Compute the internal parameters, such as the radius R of the hyperbolic disk occupied by nodes, as functions of the input parameter values, Sections 3 and 4.
2. Assign to all nodes their angular and radial coordinates on the hyperbolic plane, Section 2.2.
3. Connect each node pair by an edge with probability (the connection probability), which is a function of the coordinates of the two nodes, Sections 3 and 4.

2.2. Coordinate sampling

The assignment of node coordinates is done as follows in all the six regimes.

Angular coordinates θ of nodes are assigned by sampling them uniformly at random from interval $[0, 2\pi)$, i.e., the angular node density is uniform $\rho(\theta) = 1/(2\pi)$.

Radial coordinates $r \in [0, R]$, where R is the radius of the hyperbolic disk, are sampled from the following distribution, which is nearly exponential with exponent $\alpha > 0$,

$$\rho(r) = \alpha \frac{\sinh \alpha r}{\cosh \alpha R - 1} \approx \alpha e^{\alpha(r-R)}. \tag{1}$$

The calculation of internal parameter R is described in detail below; it is different in different regimes. Internal parameter α depends on the expected exponent γ of the power distribution $P(k) \sim k^{-\gamma}$ of nodes degrees k in generated graphs, and on the curvature of the hyperbolic space $\zeta = \sqrt{-K}$, which is set to $\zeta = 1$ by default. For temperatures $T \leq 1$, this relationship is given by

$$\gamma = 2\frac{\alpha}{\zeta} + 1, \tag{2}$$

while for $T > 1$ it becomes

$$\gamma = 2\frac{\alpha}{\zeta}T + 1. \tag{3}$$

To sample radial coordinates r according to the distribution in Eq. (1), the inverse transform sampling is used: first a random value U_i is sampled from the uniform distribution on $[0, 1]$, and then the radial coordinate of node i is set to

$$r_i = \frac{1}{\alpha} \text{acosh} \left(1 + (\cosh \alpha R - 1) U_i\right), \quad \text{for } i = 1, \ldots, N. \tag{4}$$

3. Regimes with finite $\gamma \geq 2$

3.1. $T \in (0, \infty)$: soft hyperbolic random geometric graphs

This is the most general regime in the model, from which all other regimes can be obtained as limit cases. The connection probability in this case is

$$p(x) = \frac{1}{1 + e^{\beta(\zeta/2)(x-R)}}, \tag{5}$$

where $\beta = 1/T$, and R is the radius of the hyperbolic disk occupied by nodes. The hyperbolic distance x between two nodes at polar coordinates (r, θ) and (r', θ') is given by

$$x = \frac{1}{\zeta} \text{arccosh} \left(\cosh \zeta r \cosh \zeta r' - \sinh \zeta r \sinh \zeta r' \cos \Delta\theta\right), \tag{6}$$

where $\Delta\theta = \pi - |\pi - |\theta - \theta'||$ is the angular distance between the nodes. To calculate the expected degree of a node at radial coordinate r, without loss of generality its angular coordinate can be set to zero, $\theta = 0$, so that its expected degree can be written as

$$\bar{k}(r) = \frac{N}{\pi} \int_0^R \rho(r') \int_0^\pi p(x) \, d\theta' dr'. \tag{7}$$

The expected average degree in the network is then

$$
\bar{k} = \int_0^R \rho(r)\,\bar{k}(r)dr
$$

$$
= \frac{N}{\pi} \int_0^R \rho(r) \int_0^R \rho(r') \int_0^\pi p(x)\,d\theta'\,dr'\,dr. \tag{8}
$$

Given user-specified values of input parameters N, $\beta = 1/T$, ζ and $\bar{k}$, the last equation is solved for R using the bisection method in combination with numeric evaluation of the integrals in the equation. The MISER Monte Carlo algorithm from the GSL library is used to compute the multidimensional integral in Eq. (8). The iterative bisection procedure to find R stops when the difference between the value of the computed integral in Eq. (8) and the target value of $\bar{k}$ is smaller than a threshold that is set to 10^{-2} by default.

3.2. Limit $T \to 0$: hyperbolic random geometric graphs

In the $T \to 0$ ($\beta \to \infty$) limit, the connection probability in Eq. (5) becomes

$$
p(x) = \Theta(R - x), \tag{9}
$$

where $\Theta(x)$ is the Heaviside step function, meaning that two nodes are connected if the hyperbolic distance x between them is less than R, or they are not connected otherwise. The expected average degree in the network is given by the same Eq. (8), but with $p(x)$ in the last equation. The value of R is determined using the same procedure as in Section 3.1. The only difference is that function $p(x)$ is given by Eq. (9).

3.3. Limit $T \to \infty$: soft configuration model

According to Eq. (3), in the $T \to \infty$ limit with finite α, to have finite γ, curvature should also go to infinity, $\zeta \to \infty$, such that $\eta = \zeta/T$ is finite, and instead of Eq. (3) one gets

$$
\gamma = 2\frac{\alpha}{\eta} + 1. \tag{10}
$$

More importantly, one can show that as a result of $\zeta \to \infty$, the expression for hyperbolic distance x between two nodes in Eq. (6) degenerates to

$$
x = r + r', \tag{11}
$$

meaning that in the $T \to \infty$ regime the angular coordinates are completely ignored. The connection probability becomes

$$
p(r, r') = \frac{1}{1 + e^{(\eta/2)(r+r'-R)}}, \tag{12}
$$

and the expected average degree in the network is

$$
\bar{k} = N \int_0^R \rho(r) \int_0^R \rho(r')\,p(r, r')\,dr'\,dr. \tag{13}
$$

The value of R is determined using the same combination of the bisection method and numeric integration as in the previous section, except it is applied to Eq. (13).

4. Regimes with infinite $\gamma \to \infty$

While in the $T \to \infty$ limit the angular coordinates are ignored, in the $\gamma \to \infty$ limit the radial coordinates are ignored. One can show this formally by observing that in this limit the radial node density approaches a delta function—all nodes are placed at the boundary at infinity of the hyperbolic plane, meaning that only angular coordinates determine distances between nodes.

4.1. $T \in (0, \infty)$: soft spherical random geometric graphs

In this most general case with infinite γ, one can show that the connection probability in Eq. (5) degenerates to

$$
p(\theta, \theta') = \frac{1}{1 + \lambda \left(\frac{\Delta\theta}{\pi}\right)^\beta}, \tag{14}
$$

where $\Delta\theta = \pi - |\pi - |\theta - \theta'||$ is the angular distance between the two nodes as before, while λ is a parameter controlling the average degree $\bar{k}$ in the network, analogous to R in the regimes with finite γ. Without loss of generality we can set $\theta = 0$, so that the expression for $\bar{k}$ is

$$
\bar{k} = \frac{N}{\pi} \int_0^\pi \frac{1}{1 + \lambda \left(\frac{\theta'}{\pi}\right)^\beta}\,d\theta' = N\,{}_2F_1(1, T; T + 1; -\lambda), \tag{15}
$$

where ${}_2F_1$ is the Gauss hypergeometric function, and $T = 1/\beta$. In the special case with $T = 1$, the last expression simplifies to

$$
\frac{\bar{k}}{N} = \frac{\log(1 + \lambda)}{\lambda}. \tag{16}
$$

If $T \neq 1$, the hypergeometric function in Eq. (15) cannot be evaluated using the GSL library, because the ${}_2F_1$ evaluation in the library is implemented only for the case where the fourth argument of the function ($-\lambda$ in Eq. (15)) is between -1 and 1, while for sufficiently large $N/\bar{k}$, λ is always larger than 1 in Eq. (15). To avoid this difficulty, the following transformation is used [12]:

$$
\frac{\bar{k}}{N} = {}_2F_1(1, T; T + 1; -\lambda)
$$

$$
= \frac{1}{\lambda + 1}\frac{T}{T - 1}\,{}_2F_1\left(1, 1; 2 - T; \frac{1}{\lambda + 1}\right) + \frac{1}{\lambda^T}\frac{\pi T}{\sin \pi T}. \tag{17}
$$

If $T > 1$ is an integer, the second term in (17) diverges due to the sin function in the denominator, while the first term diverges because the third parameter of the ${}_2F_1$ function is a non-positive integer. Hence, for integer values of temperature $T > 1$, their value is approximated by $T + \epsilon$, where ϵ is set to 10^{-6} by default. The error caused by this approximation is negligible. Eq. (17) (or (16) if $T = 1$) is then numerically solved for λ using the bisection method, yielding the target value of $\bar{k}$ in Eq. (15).

4.2. Limit $T \to 0$: spherical random geometric graphs

One can see from Eq. (17) that the solution for λ at small $T \ll 1$ scales with $N/\bar{k}$ as $\lambda = (N/\bar{k})^\beta$, $\beta = 1/T$. Therefore for $\beta \gg 1$ the connection probability in Eq. (14) can be written as

$$
p(\theta, \theta') = \frac{1}{1 + \left(\frac{N}{\bar{k}}\frac{\Delta\theta}{\pi}\right)^\beta}, \tag{18}
$$

which in the $\beta \to \infty$ limit becomes

$$
p(\theta, \theta') = \Theta\left(1 - \frac{N\Delta\theta}{\bar{k}\pi}\right), \tag{19}
$$

meaning that two nodes are connected if the angular distance $\Delta\theta$ between them is smaller than $\pi\bar{k}/N$,

$$
\Delta\theta < \pi\frac{\bar{k}}{N}, \tag{20}
$$

or they are not connected otherwise. This connectivity threshold ensures that the expected average degree in the network is $\bar{k}$.

Table 2
Observed properties in generated graphs (mean ± std_dev for 10^3 graph samples) with target average degree $\bar{k} = 10$.

γ＼T	0	0.5	1	2	∞
(a) Average degree.					
2	9.66 ± 2.19	9.90 ± 2.16	9.98 ± 1.08	9.88 ± 1.58	9.84 ± 1.47
3	10.08 ± 0.11	10.01 ± 0.10	10.02 ± 0.04	10.00 ± 0.08	9.99 ± 0.09
∞	10.00 ± 0.00	10.00 ± 0.00	10.00 ± 0.00	10.00 ± 0.00	10.00 ± 0.00
(b) Average clustering.					
2	0.88 ± 0.00	0.75 ± 0.00	0.29 ± 0.00	0.36 ± 0.01	0.36 ± 0.01
3	0.79 ± 0.00	0.41 ± 0.00	0.42 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
∞	0.75 ± 0.00	0.33 ± 0.00	0.30 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

(a) $\gamma = 2$.

(b) $\gamma = 3$.
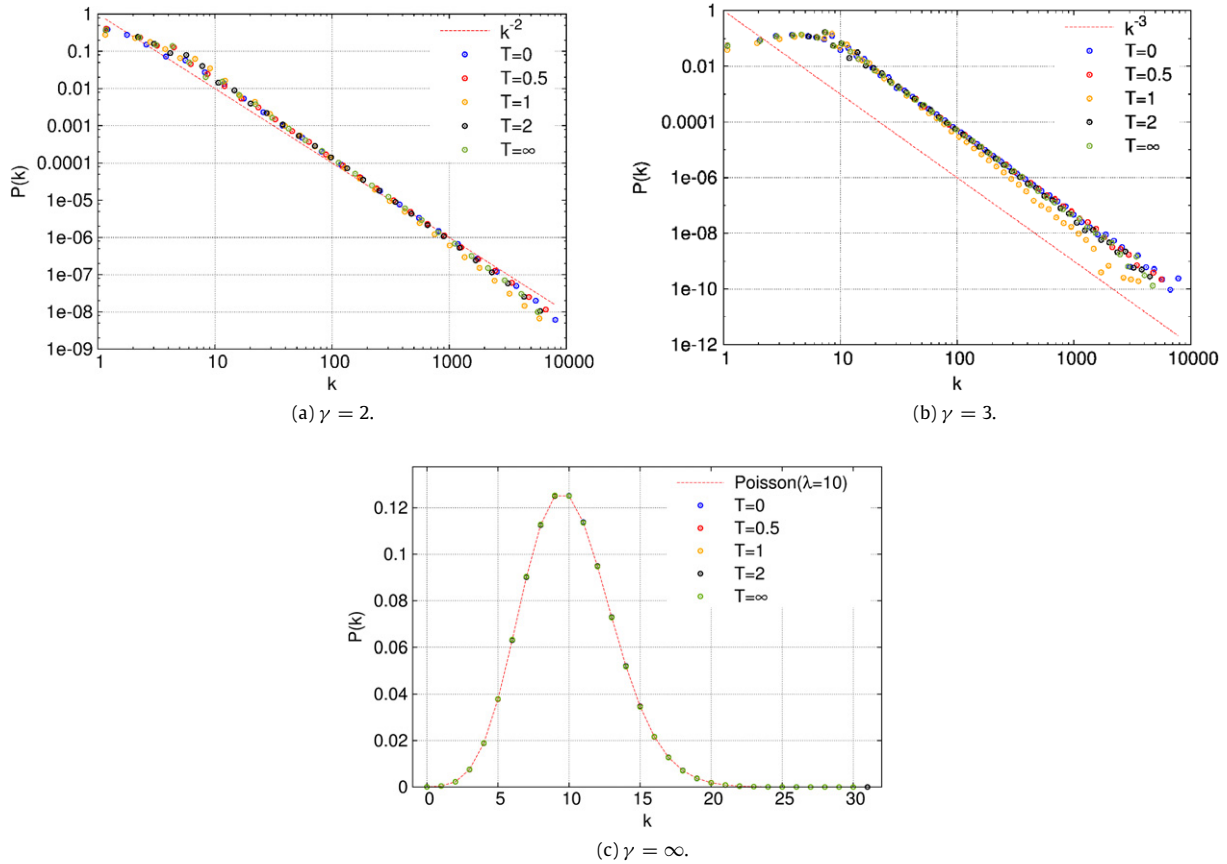

(c) $\gamma = \infty$.

Fig. 1. Observed degree distributions in generated networks of size $N = 10^4$ and target average degree $\bar{k} = 10$. The results are averaged across 10^3 generated graphs for each combinations of the parameters.

4.3. Limit $T \to \infty$: Erdős–Rényi graphs

In this most degenerate regime, both angular and radial coordinates are completely ignored. This regime is formally achieved by keeping both α and ζ finite while letting $T \to \infty$. One can then show that the connection probability in Eq. (5) degenerates to

$$p(x) = \frac{1}{1 + \frac{N}{\bar{k}}}, \tag{21}$$

which for sparse graphs with $\bar{k} \ll N$ tends to $p(x) = \bar{k}/N$, i.e., the connection probability in classical (Erdős–Rényi) random graphs.

5. Simulations

Tables 2(a) and (b) show, respectively, the average degree and clustering values in generated graphs for different regimes, 10^3 samples in each regime. All the regimes match the target average degree $\bar{k} = 10$, although low values of γ lead to much higher fluctuations. For any $\gamma > 2$, average clustering decreases with temperature from a maximum at $T = 0$ to zero at $T \to \infty$.

Fig. 1 shows the observed distribution of node degrees for three different values of γ. As expected, for finite γ, distributions follow a power-law $P(k) \sim k^{-\gamma}$. In the limit at $\gamma \to \infty$ the observed degree distributions follow exactly the Poisson distribution with mean $\mu = \bar{k}$.

Fig. 2 shows clustering in generated graphs. For finite γ, low-degree nodes have stronger clustering than high-degree nodes. In the case of $\gamma \to \infty$, all nodes have similar clustering values.

(a) $\gamma = 2$.

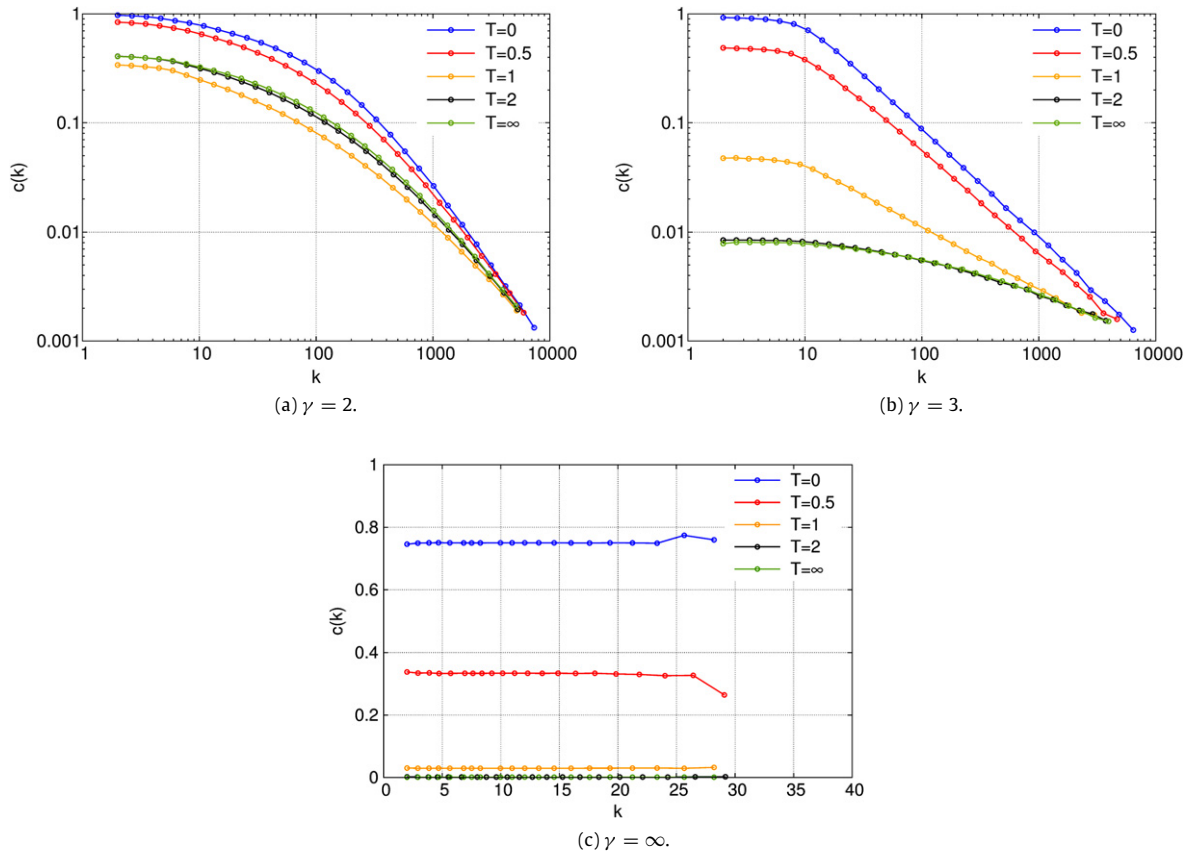(b) $\gamma = 3$.

(c) $\gamma = \infty$.

Fig. 2. Observed clustering in generated networks of size $N = 10^4$ and target average degree $\bar{k} = 10$. The results are averaged across 10^3 generated graphs for each combinations of the parameters.

References

[1] S.N. Dorogovtsev, Lectures on Complex Networks, vol. 24, Oxford University Press, Oxford, 2010.

[2] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.

[3] M. Newman, Networks: An Introduction, Oxford University Press, Oxford, 2010.

[4] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, M. Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E 82 (2010) 036106.

[5] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, D. Krioukov, Popularity versus similarity in growing networks, Nature 489 (2012) 537–540.

[6] M. Boguñá, F. Papadopoulos, D. Krioukov, Sustaining the internet with hyperbolic mapping, Nat. Comm. 1 (2010) 62.

[7] F. Papadopoulos, C. Psomas, D. Krioukov, Network mapping by replaying hyperbolic growth, IEEE/ACM T. Netw. 23 (2014) 198–211.

[8] J. Park, M. Newman, Statistical mechanics of networks, Phys. Rev. E 70 (2004) 066117.

[9] M.D. Penrose, Random Geometric Graphs, vol. 5, Oxford University Press, Oxford, 2003.

[10] M.D. Penrose, Connectivity of soft random geometric graphs, arXiv:1311.3897.

[11] P. Erdős, A. Rényi, On random graphs, Pub. Math. Deb. 6 (1959) 290–297.

[12] R.C. Forrey, Computing the hypergeometric function, J. Comput. Phys. 137 (1997) 79–100.