1. (a) We construct 2-universal hash function h_1, h_2 as follows:

$h_1(x) = x \bmod 2$

$h_2(x) = (x+1) \bmod 2$

To clarif, suppose distinct $a, b \in \{1, 2, 3, 4\}$

$Pr[h_1(a) = h_1(b)] = Pr[a \bmod 2 = b \bmod 2] = Pr[(a-b) \bmod 2 = 0]$

As $a \neq b$, so pair (a, b) will be 6 choices.

$(1,2)$ $(1,3)$ $(1,4)$ $(2,3)$ $(2,4)$ $(3,4)$

$(1-2) \bmod 2 = 1$ $(1-3) \bmod 2 = 0$ $(1-4) \bmod 2 = 1$ $(2-3) \bmod 2 = 1$ $(2-4) \bmod 2 = 0$

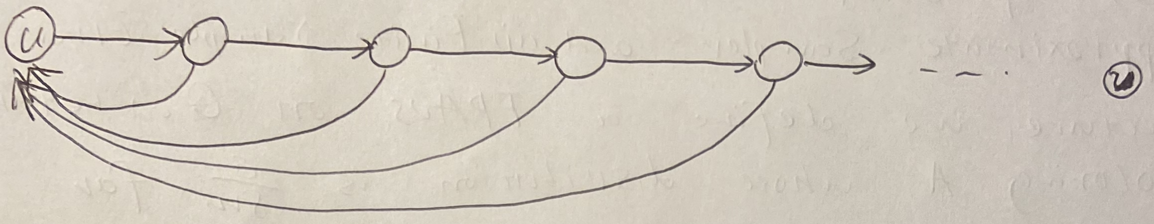
$(3-4) \bmod 2 = 1$

so $Pr[(a-b) \bmod 2 = 0] = \frac{1}{2}$

The analysis for h_2 is same as above

We can prove h_1, h_2 are 2-universal

(b) We can construct a directed graph like this



As we can see, each node will have a path directed to the starting node u. In this example, the expected ~~[illegible]~~ steps from u to v is $2^{\Omega(n)}$.
Because each step we have chance to go back to start point u, so the $Pr[u \to v] = (\frac{1}{2})^n$.
So the steps that we get v is $\cancel{[illegible]}^r \geq 2^{\Omega(n)}$

~~[illegible]~~

(c) In k-colorings problem, there are two sources of error

(i) Approximate Sampler and (ii) Finite Sample Error

In lecture, we define a FPAUS on G_{i-1} return a coloring A whose distribution is $\frac{\epsilon}{8m}$ - far from uniform in TV distance

$$\left\| \Pr[A \text{ colors } G_i] - \frac{C_i}{C_{i-1}} \right\|_{TV} \le \frac{\epsilon}{8m}$$

Here each sample is involded into C_{i-1}/C_i. ~~with~~ w.h.p.

In other way

$$\| w(x) - U(x) \|_{TV} \le \frac{\epsilon}{8m}$$

which $w(x)$ means ~~sample~~ select a sample from distribution $w(x)$, $U(x)$ means select a sample from C_{i-1} uniformly.

So from this r coloring from G_{i-1}, let Z_i be how many property color G_i, From above:

$$\frac{rC_i}{C_{i-1}}\left(1-\frac{\epsilon}{6m}\right) \le r\cdot\left(\frac{C_i}{C_{i-1}} - \frac{\epsilon}{8m}\right) \le \mathbb{E}[Z_i] \le r\cdot\left(\frac{C_i}{C_{i-1}} + \frac{\epsilon}{8m}\right) \le \frac{rC_i}{C_{i-1}}\left(1+\frac{\epsilon}{6m}\right)$$

Then as lectures we can get ~~approximating~~ approximate counting of k-colorings done in class.

2. (a) Because this markov chain has self-loop, which means $S_{t+1} = S_t$ w.p. $\frac{1}{2}$, so it must be aperiodic. As every time the markov chain will add a element from $[n]/S_t$ and delete an element from S_t, w.p. $\frac{1}{2}$, every subset with k elements will be reachable in this markov chain, which means $P^t_{x,y} > 0$. So this MC is irreducible.

(b) Suppose a state x has neighbors $N(x)$, which are all states ~~which has~~ that have one element from $[n]/s$ different with it. So this state has $(n-k) \times n$ edges, in other way, the degree of x $\deg(x) = (n-k) \times n$ is a constant. As $\pi_i = \frac{\deg(i)}{2|E|}$

$$\pi_x = \frac{(n-k) \times n}{2|E|} = \frac{1}{C_n^k} = \text{constant}$$

Consequently, the stationary distribution is uniform over Ω

(c) As we can see, X_t is satisfied original Markov Chain. which means

$Pr[X_{t+1} = x' \mid Z_t = (x, y)] = Pr[X_{t+1} = x' \mid X_t = x]$ satisfies MC S_t

As for Y_t, we can find that

$Pr[Y_{t+1} = y \mid Z_t = (x, y)$, choose first part$]$

$= Pr[Y_{t+1} = y \mid Y_t = y$, chose first part$] = \frac{1}{2}$

$Pr[Y_{t+1} = y' \mid Z_t = (x, y)$, choose second part$]$

$= \cancel{Pr[i \in X_t \cap Y_t]} \cancel{Pr[i \in X_t \cap Y_t]} = Pr[$choose i'; choose $j']$

$= \frac{1}{k} \cdot \frac{1}{(n-k)} = Pr[Y_{t+1} = y' \mid Y_t = y] = Pr[X_{t+1} = x' \mid X_t = x]$

so individually, each X_t and Y_t ~~are~~ a satisfied Original Markov Chain (S_t).

Consequently, (X_t, Y_t) is a valid couping for the MC S_t

(d) Because $\mathbb{E}[q_{t+1} \mid q_t] \le (1-\frac{1}{k}) q_t$, use Markov's Inequality

$$\Pr[q_t \geq 1] \le \mathbb{E}[q_t] \le k(1-\frac{1}{k})^t$$

with $|q_t| \le k$. When $q_t = 0$, $X_t = \xi_t$, so coupled.

Let T_0 be the first time q_t becomes 0. The above implies $\mathbb{E}[T_0] = O(k\log k)$

Use Markov's Inequality again, let $\mathbb{E}[T_0] = ck\log k$

$\cancel{\Pr[T \geq 3ck\log k] \le \frac{O(k\log k)}{ck\log k}}$

c is a constant.

$$\Pr[T_0 \geq 3ck\log k] \le \frac{\mathbb{E}[T_0]}{3ck\log k} = \frac{1}{3}$$

$t_{min}(\frac{1}{3}) \le O(k\log k)$

Using ~~lemma~~ Theorem 12.6 in MU Book

$t_{min}(\epsilon) = \cancel{O(k\log k} \quad O(k\log(k/\epsilon))$

(e) When (i), $q_{t+1} - q_t = -2$

(ii) $q_{t+1} - q_t = -2$

(iii) $q_{t+1} - q_t = -2$

(iv) $q_{t+1} - q_t = 0$

(f) As we can see in (d)
q_t only decrease when in "otherwise" part case (i), (ii), (iii)
Thus

$$\mathbb{E}[q_{t+1} - q_t \mid D_t] = \mathbb{E}[q_{t+1} - q_t \mid D_t, \overset{\text{in cases}}{\cancel{\text{choose}}}\ \cancel{\text{"otherwise"}}\ (i), (ii), (iii)]$$

$$\Pr[\overset{\text{cases}}{\cancel{\text{choose}}}\ (i)(ii)\ (iii)]$$

$$= (-2) \times \left(\Pr[\overset{\text{case}}{\cancel{\text{choose}}}\ (i) + \overset{\text{case}}{\cancel{\text{choose}}}\ (ii) + \overset{\text{case}}{\cancel{\text{choose}}}\ (iii)]\right) \times \frac{1}{2}$$

$$= (-2) \times \Pr[\overset{\text{case}}{\cancel{\text{choose}}}\ (i)] + (-2) \times \Pr[\overset{\text{case}}{\cancel{\text{choose}}}\ (ii)] + (-2) \times \Pr[\overset{\text{case}}{\cancel{\text{choose}}}\ (iii)]$$

$$= (-2) \times \frac{1}{2}\left(\frac{(k - \frac{q_t}{2})(\frac{q_t}{2})}{k(n-k)} + \frac{\frac{q_t}{2}(n-k-(k-\frac{q_t}{2}))}{k(n-k)} + \frac{\frac{q_t}{2}(\frac{q_t}{2}-1)}{k(n-k)}\right)$$

$$= -\frac{n-2}{k(n-k)} q_t$$

~~so~~ with $k \geq 2$

$$-\frac{n-2}{k(n-k)} q_t \leq -\frac{1}{k} q_t$$

So ~~$\mathbb{E}[q_{t+1}\ q_{t+1}$~~ $\mathbb{E}[q_{t+1} - q_t \mid D_t] = -\frac{n-2}{k(n-k)} q_t \leq -\frac{1}{k} q_t$

Q.E.D.

3.(a) It's easily to show that r_i's ~~...~~ expectation $\mathbb{E}[|r_i|] < \infty$, becauce it's just a ratio.

Now, to show $\mathbb{E}[r_{i+1} | R_0, R_1 \dots R_i] = r_i$

$$\mathbb{E}[r_{i+1} | R_0, R_1, \dots R_i] = \mathbb{E}\left(\frac{R_{i+1}}{R+G-i-1} \,\middle|\, R_0, R_1, \dots, R_i\right)$$

$$= \frac{1}{R+G-i-1} \mathbb{E}[R_{i+1} | R_0, R_1, \dots, R_i)$$

$$= \frac{1}{R+G-i-1} \left[(R_i - 1) Pr[\text{pick red ball}] + R_i Pr[\text{pick green ball}]\right)$$

$$= \frac{1}{R+G-i-1} \left[(R_i - 1)\left(\frac{R_i}{R+G-i}\right) + R_i\left(\frac{R+G-i-R_i}{R+G-i}\right)\right]$$

$$= \frac{R_i}{R+G-i} = r_i$$

As above, we demonsterate. the sequence $r_0, \dots, r_n$ forms a martingale w.r.t. $R_0, \dots, R_n$

(b) Here we first prove

$$|r_k - r_{k-1}| = \left| \frac{R_k}{R+G-k} - \frac{R_{k-1}}{R+G-(k-1)} \right| \le |C_k| = 1$$

$$\left| \frac{R_k}{R+G-k} - \frac{R_{k-1}}{R+G-(k-1)} \right|$$

$$= \left| \frac{(R+G)(R_k - R_{k-1}) + kR_{k-1} - (k-1)R_k}{(R+G)^2 - (2k-1)(R+G) + k(k-1)} \right|$$

$$\le \frac{R+G+R_k}{(R+G-k)(R+G-k+1)} \le 1 = |C_k|$$

with Azume-Hoeffding Inequality, and

$$\mathbb{E}[r_n] = \frac{R}{R+G}$$

$$Pr\left[\left|r_n - \frac{R}{R+G}\right| > \lambda\right] \le 2e^{-\frac{\lambda^2}{2n}}$$

let $\lambda = \sqrt{2n\ln(\frac{2}{\delta})}$ ~~which δ is in $(0, \frac{1}{2})$ is a~~

which δ is a small number in $(0,1)$

$$Pr\left[\left|r_n - \frac{R}{R+G}\right| > \sqrt{2n\ln(\tfrac{2}{\delta})}\right] \le \delta$$

so w.h.p $(1-\delta)$, $\frac{R}{R+G} - \sqrt{2n\ln(\frac{2}{\delta})} < r_n$

$< \frac{R}{R+G} + \sqrt{2n\ln(\frac{2}{\delta})}$

r_n will concentrates around $\frac{R}{R+G}$

as $R_n = n r_n$

R_n = # of red balls removed will concentrates on

$$\frac{nR}{R+G}$$

Q.E.D.